TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

GEORGE DAVID BIRKHOFF

ARTHUR BYRON COBLE

LUTHER PFAHLER EISENHART

WITH THE COÖPERATION OF

OLIVER E. GLENN
EDWARD KASNER
HOWARD H. MITCHELL
FOREST R. MOULTON
WALLIE A. HURWITZ
AUBREY J. KEMPNER
CHARLES N. MOORE
FRANCIS R. SHARPE
ERNEST J. WILCZYNSKI
DUNHAM JACKSON
WILLIAM R. LONGLEY
ROBERT L. MOORE
J. H. M. WEDDERBURN

VOLUME 23
1922

PUBLISHED BY THE SOCIETY
LANCASTER, PA., AND NEW YORK
1922-3

RELATIVELY UNIFORM CONVERGENCE AND THE CLASSIFICATION OF FUNCTIONS*

BY

E. W. CHITTENDEN

INTRODUCTION

The classification of functions of real variables given by Baire† is based upon iteration of simple limits. The classes of functions obtained are found to be closed relative to uniform convergence. Several types of convergence which are more exacting than simple convergence and less exacting (in general) than uniform convergence have proved useful in analysis, notably the sub-uniform convergence of Arzelà,‡ the uniform convergence in general of Weyl,§ and the relatively uniform convergence of E. H. Moore.|| It is of interest to inquire into the relation of these types of convergence to the Baire classification of functions. The present paper is concerned with the problem presented by the relatively uniform convergence of Moore which will be spoken of as convergence (R).

* Presented to the Society, September 8, 1920.

† Annali di Matematica, 1900.

‡ Arzelà, *Sulle serie di funzioni*, Memorie della Reale Accademia degli Scienze di Bologna, ser. 5, vol. 8 (1900). This type of convergence is called "convergenza uniforme a tratti" by Arzelà. The designation "sub-uniform" appears to have originated with E. H. Moore, Bulletin of the American Mathematical Society, vol. 7, p. 257.

§ H. Weyl, Mathematische Annalen, vol. 67 (1909), p. 225. Weyl has introduced under the designation, uniform convergence in general, the following type of convergence. Let a sequence of functions converge on a measurable set P. Then there is a set Q whose measure is as near the measure of P as you please on which the convergence is uniform.

If $e_1, e_2, \ldots, e_k, \ldots$ is a monotonic sequence of positive numbers with the limit zero, there will be sets

$$Q_1, Q_2, \ldots, Q_k, \ldots$$

such that mes $(P - Q_k) \leqq e_k$, Q_{k+1} contains Q_k, and the convergence is uniform on each set Q_k. If $Q = \lim Q_k$ then $P - Q$ is of measure zero. It follows from the theorem of § 14 below that uniform convergence in general is equivalent to relatively uniform convergence excepting a set of points of measure null.

|| *Introduction to a form of general analysis, The New Haven Mathematical Colloquium*, Yale University Press, New Haven, 1910; p. 30, § 7d.

We are indebted to Baire for a detailed investigation of properties of functions of the classes 1, 2, 3,* but the general theory of functions of class α, where α denotes a transfinite number with at most a denumerable infinity of predecessors, was developed principally by Lebesgue, who was able, incidentally, to demonstrate the existence of functions of every class.†

Ch. J. de la Vallée-Poussin has given a very clear and concise account of the results obtained by Baire and Lebesgue in his monograph entitled *Intégrales de Lebesgue, Fonctions d'Ensemble, Classes de Baire.*‡ He has introduced some characteristic improvements in the exposition involving the discovery of new and powerful theorems, and exhibited the important rôle of the sets which are at once O and F of class α (that is ambiguous (A) of class α). The reader is referred to this monograph for the proofs of the theorems relating to the classification of Baire which are assumed in the following discussion. The notation and terminology of Vallée-Poussin are employed in this paper.

In previous papers the writer has shown that there exist sequences of functions which converge but do not converge uniformly relative to any scale function,§ and has found necessary and sufficient conditions that a function be the limit of a sequence of *continuous* functions convergent (R),|| thus completely characterizing the functions of what may be called the class R_1.

The present paper presents the results of further investigations on this subject. Let f denote a function of class α of Baire. If for every value of the real constant a the sets $(f > a)$, $(f < a)$ are ambiguous (A) of class α, the function f is said to be of class A_α, a proper subclass of class α. Denoting by R_α the class of functions defined in terms of convergence (R) as class α of Baire is defined in terms of simple convergence, then all functions of the class R_α which are not of class $<\alpha$ in the classification of Baire are contained in the class A_α. There is a subclass D_α of R_α (see §5) such that every function of class α is the limit of a *uniformly* convergent sequence of functions of D_α. If $\alpha = 1$, $R_\alpha = A_\alpha$. There are indications that the conditions which functions of the class R_α $(\alpha > 1)$ must satisfy are stronger than those imposed on the functions of A_α, but no examples demonstrating this are available.

A number of theorems dealing with the general theory of functions of class α have been found. These theorems are presented as they naturally occur in the development of the subject.

While the discussion is confined to functions defined on limited perfect sets in space of n dimensions this restriction is not essential, and it is possible to apply

* Acta Mathematica, vol. 30 (1906).
† Journal de Mathématiques, ser. 6, vol. 1 (1905).
‡ Gauthier-Villars, Paris, 1916.
§ These Transactions, vol. 15 (1914), pp. 197–201.
|| These Transactions, vol. 20 (1919), pp. 179–184.

the theory with suitable and fairly obvious modifications to classes of functions on a great variety of ranges. As an illustration we cite an article by the writer, *On a generalization of a theorem of Baire.**

I. Classification of functions and sets of points

1. **The Baire classification of functions.** The functions which enter the following discussion are assumed to be defined at every point of a fundamental limited perfect set P in space of n dimensions. The functional values are limited to the class of real numbers; that is, the functions are *finite*, completely defined, single-valued, real-valued functions of the variable point M with the range P.

Baire's classification of functions is defined as follows: the continuous functions are of class zero (B_0); the discontinuous functions which are limits of sequences of continuous functions are of class 1 (B_1), etc.; and in general the functions of class α (B_α) are those limit functions of denumerable sequences of functions of classes preceding class α which are not functions of class $< \alpha$.

2. **Lebesgue's classification of sets of points.** Lebesgue has introduced a classification of sets of points based on Baire's classification of functions which has proved a remarkably effective tool in the investigation of the properties of the classes B_α. A set E is said to be *open* or O of class α if there is a function Θ of class $\leqq \alpha$ such that E is the set of all points of P at which $\Theta > 0$. This set E is represented by the notation $E = (\Theta > 0)$. A set E is *closed* or F of class α if there is a function Θ of class $\leqq \alpha$ for which E is the set $E = (\Theta = 0)$.

Vallée-Poussin has called attention to the importance of the sets which are at once O and F of class α and has called them *ambiguous* or A of class α.

The proofs of the following statements regarding sets O, F, and A of class α will be found in the monograph of Vallée-Poussin already cited, §§ 132–134, pp. 132–139.

Every set O of class α is F of class $\alpha + 1$. Every set F of class α is O of class $\alpha + 1$. Therefore, every set which is either O or F of class α is A of class $\alpha + 1$.

We denote by Σ the operation of taking the sum of a finite or denumerably infinite sequence of sets, by Π the corresponding product. Then if O_α, F_α, A_α, signify sets which are O of class α, etc., we have

$$\sum O_\alpha = O_{\alpha+1}, \quad \prod F_\alpha = F_{\alpha+1},$$

and

$$\sum A_\alpha = A_{\alpha+1}, \quad \prod A_\alpha = A_{\alpha+1}.$$

---

*Bulletin of the American Mathematical Society, vol. 27 (1920–21), p. 5.

3. The limit of a sequence of sets of points. Given a sequence $E_1, E_2, E_3, \ldots, E_n, \ldots$ of sets of points we define the greatest limit of the sequence by the formula:

$$\overline{\lim}\, E_n = (E_1 + E_2 + E_3 + \cdots)(E_2 + E_3 + \cdots)(E_3 + \cdots)\cdots,$$

and the least limit by

$$\underline{\lim}\, E_n = E_1E_2E_3\cdots + E_2E_3\cdots + E_3\cdots + \cdots.$$

If both formulas give the same result, $\lim E_n$ exists. (Cf. Vallée-Poussin, loc. cit. §10, pp. 8, 9.)

It follows that if the sets E_n are O of class α, $\overline{\lim}\, E_n$ is of the form a product of sums of sets which are O of class α. In the notation of §2 we have

$$\overline{\lim}\, E_n = \prod \sum O_\alpha = \prod O_\alpha = F_{\alpha+1}.$$

That is, $\overline{\lim}\, E_n$ is F of class $\alpha + 1$. Under the same hypothesis

$$\underline{\lim}\, E_n = \sum \prod O_\alpha = \sum F_{\alpha+1} = O_{\alpha+2}.$$

If the E_n are F of class α, similar reasoning shows that $\overline{\lim}\, E_n$ is F of class $\alpha + 2$ while $\underline{\lim}\, E_n$ is O of class $\alpha + 1$.

Consequently, if the E_n are ambiguous (A) of class α, $\overline{\lim}\, E_n$ is F of class $\alpha + 1$, $\underline{\lim}\, E_n$ is O of class $\alpha + 1$, and $\lim E_n$ when it exists is both O and F of class $\alpha + 1$, that is A of class $\alpha + 1$.

The following important theorem is readily inferred from properties of sets A of class α given by Vallée-Poussin (loc. cit., pp. 136–8); if $\alpha > 0$, every set A of class α is a limit of sets A of class $< \alpha$ and every set which is a limit of sets A of class $< \alpha$ is A of class α.

4. The characteristic function of a set of points. The function φ equal to one at the points of E and to zero elsewhere is called by Vallée-Poussin the *characteristic of* E. If the characteristic function φ is of class α, then E is A of class α. Conversely, if E is A of class α then φ is of class $\leqq \alpha$. If a set E is O or F of class α its characteristic is of class $\leqq \alpha + 1$.

If Q_1 and Q_2 have characteristics φ_1, φ_2, the characteristic of the product Q_1Q_2 is $\varphi_1\varphi_2$ while the characteristic of the sum $Q_1 + Q_2$ is given by the formula

$$\varphi_1 + \varphi_2 - \varphi_1\varphi_2.$$

This reduces to $\varphi_1 + \varphi_2$ when Q_1, Q_2 have no common element, that is, do not overlap. It follows that if φ_1, φ_2 are of class $\leqq \alpha$ then $Q_1 + Q_2$ is A of class α.

5. Relations between the classification of functions and the classification of sets of points. Let $(a < f < b)$, $(a \leqq f \leqq b)$ denote the sets of all points at which the respective inequalities hold for a given function f. Then if f is of class α, the set $(a < f < b)$ is O of class α, and the set $(a \leqq f \leqq b)$ is F of class α for every pair of values of a and b $(> a)$. Furthermore there exist values of a and b such that the set $(a < f < b)$ is not O of any class $< \alpha$.

The functions of class α for which the set $(\alpha < f < b)$ is always A of class α form an important subclass of the functions of class α denoted by the symbol A_α.

If a set Q is A of class α its complement $P - Q$ is also A of class α. It follows at once that a function of class α which assumes only a finite number of distinct values belongs to the class A_α, in fact to a subclass of A_α, which we denote by F_α.

If the values of a function f of class α form at most a denumerably infinite set $a_1, a_2, \ldots, a_n, \ldots$, which has no finite limit point, the function belongs to the class A_α. We denote by the symbol D_α the class of all such functions. The class D_α contains the class F_α.

II. On the composition and decomposition of sets of points in relation to the classification of Lebesgue

6. A theorem on the decompositon of sets of class $\alpha > 0$.

THEOREM. *If a set Q of class α is the sum of sets Q_1, Q_2 of class α which do not overlap there will exist sequences of sets K_n, K_{1n}, K_{2n} of class $< \alpha$ such that*

$$K_n = K_n + K_{2n}(n)$$

and $\lim K_n = Q$, $\lim K_{1n} = Q_1$, $\lim K_{2n} = Q_2$.

Since Q and Q_1 are of class α there exist sequences of sets H_n, H_{1n} of class $< \alpha$ which have the respective limits Q, Q_1. Set

$$K_n = H_n, \quad K_{1n} = H_n H_{1n}, \quad K_{2n} = K_n - K_{1n}.$$

Then $K_n = K_{1n} + K_{2n}$ and the sets K_{1n}, K_{2n} are of class $< \alpha$ (n). If M is a point of Q_1 it is a point of Q and there will exist n_1 such that for all $n \geqq n_1$, M will belong to H_n and H_{1n}, therefore to K_{1n}. Consequently $\lim K_{1n} = Q_1$. If M is a point of Q_2 it is for all $n \geqq n_2$ an element of H_n and the complement of H_{1n}. Therefore M belongs to $H_n (P - H_{1n}) = K_n(P - H_{1n}) = K_n - K_{1n} = K_{2n}$. Hence, $\lim K_{2n} = Q_2$. This theorem may be generalized in an obvious way.

Let φ_{1n}, φ_{2n} be the characteristics of K_{1n}, K_{2n}. The function $\varphi_n = a_1\varphi_{1n} + a_2\varphi_{2n}$ is of class $< \alpha$ and converges with increasing n to the function $\varphi = a_1\varphi_1 + a_2\varphi_2$ of class α where φ_1 and φ_2 are the respective characteristics of Q_1, Q_2.

7. Theorems on decomposition.

THEOREM I. *Let $E_1, E_2, \ldots, E_n$, beany system of sets O of class $\alpha > 0$ such that*

$$P = E_1 + E_2 + \cdots + E_n.$$

Then there exists a normal system $H_1, H_2, \ldots, H_n$ of sets A of class α such that:

(1) *the sets, $H_1, H_2, \ldots, H_n$ do not overlap*;

(2) *the set* H_i *is contained in* E_i $(i = 1, 2, \ldots, n)$;

(3) $P = \sum_{i=1}^{n} H_i$.

Since each set E_i is O of class α it may be represented in the form $E_i = \sum_{m=1}^{\infty} F_{im}$ where F_{im} is A of class α (m) (cf. Vallée-Poussin, loc. cit., p. 138, 5°), and we may assume without loss of generality that $F_{i,m+1}$ contains F_{im}. Set $H_{11} = F_{11}$, $H_{i0} = [0]$, and

$$H_{im} = F_{im} P\Big(- \sum_{j=1}^{j=i-1} H_{jm} - \sum_{j=i+1}^{j=n} H_{j,m-1} \Big).$$

Then H_{im} is A of class $\alpha(im)$. We wish to establish the following properties of the system $((H_{im}))$:

(A) H_{im+1} contains H_{im} (im);

(B) if i and j are unequal, H_{im} and H_{jn} have no element in common for any pair of values of m and n;

(C) every element of P is contained in some class H_{im}.

It is clear that $H_{i1} = F_{i1}(P - \sum_{j=1}^{i-1} F_{j1})$, and that in consequence the H_{j1} do not overlap. Since $H_{12} = F_{12}(P - \sum_{j=2}^{n} H_{j1})$ it is evident that H_{12} contains H_{11}, and it is likewise easy to see that H_{i2} contains H_{i1} and that the H_{i2} do not overlap. It is now simply a matter of induction to establish property (A) and to show that for fixed m the H_{im} do not overlap. To establish the remaining part of (B), we observe that if m is greater than n, then H_{jm} contains H_{jn} and has no element in common with H_{im}. Finally, by hypothesis, every element of P lies in some E_i and therefore in some F_{im}. If this element does not belong to H_{im}, it is because it belongs to one of the classes $H_{jm}(j < i)$ or $H_{j,m-1}(j > i)$.

We now set $H_i = \sum_{m=1}^{\infty} H_{im}$. It follows from the properties of the system $((H_{im}))$ just established that the system (H_i) satisfies the conditions (1, 2, 3) of the theorem. But H_i is certainly O of class α and its complement is therefore F of class α. But the complement of H_i is $\sum_{j=1}^{i-1} H_j + \sum_{j=i+1}^{n} H_j$ and is consequently O of class α. It follows that H_i is A of class $\alpha(i)$. This completes the proof of the theorem.

THEOREM II. *If* $P = \sum_{n=1}^{\infty} E_n$, E_n *is* O *of class* α *for every* n, *and* $\prod_{n=1}^{\infty} E_n = [0]$, *there exists a sequence of classes* F_n, A *of class* α (n), *such that* $F_n F_{n'} = [0]$ $(n' \neq n)$, $\sum_{n=1}^{\infty} F_n = P$, *and* $F_n \leqq E_n(n)$.

From the preceding theorem we have, if we denote $\sum_{m=n+1}^{\infty} E_m$ by R_n, a decomposition of

$$P = E_1 + E_2 + \cdots + E_n + R_n$$

into a normal system:

$$P = F_{n1} + F_{n2} + \cdots + F_{nn} + G_n.$$

Then if we set

$$F_1 = F_{11},$$
$$F_2 = (F_{12} + F_{22})\ (P - F_1),$$

and in general,

$$F_n = \Big(\sum_{i=1}^{n} F_{in}\Big)\Big(P - \sum_{i=1}^{n-1} F_i\Big),$$

we have the desired sequence $\{F_n\}$.

III. FUNCTIONS OF CLASS α — REDUCTION — SUPERPOSITION

8. A generalization of a theorem of Vallée-Poussin. The following theorem is a direct generalization of a theorem stated by Vallée-Poussin for functions of class 1.

THEOREM. *Every function of class* $\alpha > 0$ *is the limit of a uniformly convergent sequence of functions of class* D_α.

In fact if we let $E_n = (\overline{n-1}e < f < \overline{n+1}e)$, where e denotes an arbitrarily small positive number, the set E_n is O of class α and $P = \sum_{n=-\infty}^{+\infty} E_n$. Applying the results of the preceding article we obtain a sequence of sets F_n, A of class α, which do not overlap, such that F_n is contained in E_n and $\sum_{n=-\infty}^{+\infty} F_n = P$. If we set $\varphi = ne$ on $F_n(n)$, φ is defined on P, is of class D_α, and it is evident from the inclusion of F_n in E_n that $|\varphi - f| \leqq e$.

The theorem follows immediately.

9. A theorem on superposition of functions of class A_α.

THEOREM. *If* $Q_1, Q_2, \ldots, Q_m, \ldots$ *is a sequence of non-overlapping sets whose characteristics* φ_m *are of class* $\leqq \alpha$, $f_1, f_2, \ldots, f_m, \ldots$, *is a sequence of*

functions of class A_α at most, and $P = \sum_{m=1}^{\infty} Q_m$, then

$$f = \sum_{m=1}^{\infty} \varphi_m f_m$$

is of class A_α on P.

Consider the set $E = (a \leqq f \leqq b)$. It is of the form $E = \sum_{m=1}^{\infty} U_m$, where $U_m = (a \leqq \varphi_m f_m \leqq b)$ and is A of class α (m). But $Q_m - U_m$ is A of class $\alpha(m)$ and therefore both $P - E = \sum_{m=1}^{\infty} (Q_m - U_m)$ and E are O of class α. It follows that E is A of class α, which was to be proved.

10. Reduction-functions of class α on a subset of P. If a function f is defined on P its reduction $f(Q)$ relative to a subset Q of P is the function on Q which has the functional values of f.*

We shall find it convenient to classify functions on an arbitrary subset Q of the fundamental set.

A function f is of class α on Q in case it is not of class $< \alpha$ on Q and is the reduction relative to Q of a function ψ which is the limit on P of a sequence of functions of class $\leqq \alpha$ on P converging uniformly on Q.

It is evident that the reduction relative to Q of a function of class α on P is of class $\leqq \alpha$ on Q, but that a function of class α on Q may not be the reduction of a function of class α, although it must be the reduction of a function of class $\alpha + 1$. It will be noticed that the reduction of a function of class α may be of lower class; for example, the reduction relative to Q of the characteristic of Q is of class zero. Not every continuous function on an open set is of class zero.† For example, let Q denote the sequence $1, \frac{1}{2}, \cdots, \frac{1}{n}, \cdots$, and let $f(\frac{1}{n})$ equal 1 (n odd), 2 (n even). Then $f(\frac{1}{n})$ is of class 1, although continuous on Q.

It follows readily from a theorem of Vallée-Poussin (loc. cit., § 130, p. 130) that the class α of functions on Q as defined above has the usual closure relative to uniform convergence. The functions of class $\alpha + 1$ whose reductions relative to Q are of class α on Q form a subclass of the class $\alpha + 1$.

The definition of function of class α on Q given above is to be compared with the following definition of Vallée-Poussin (loc. cit., § 135, p. 139): *A function f is within e of class $\alpha > 0$ on a set Q if, on Q, f is within e of being a limit of functions of class $< \alpha$ on P.*

The class of functions, within e of class α on Q for every $e > 0$, contains the functions of class α on Q as defined above. If Q is a closed set these two classes

* E. H. Moore, loc. cit., p. 91, § 52.

† It is easy to see that every function which is continuous on an arbitrary set Q is the reduction of a semi-continuous function, that is, a function of class A_1.

of functions are the same for every value of $\alpha > 0$. It would be of interest to know whether there exist classes Q for which this conclusion fails to hold.

11. THEOREM. *Let Q be an arbitrary subset of P and suppose that $f_1, f_2, \ldots, f_n \ldots$, is a sequence of functions of class $\leqq \alpha$ converging uniformly on Q. Then Q is contained in a set $\overline{Q}$ which is F of class α and the sequence converges uniformly on $\overline{Q}$.*

Because of the uniform convergence on Q we have for every integer m and integer n_m such that for all $n \geqq n_m$ the inequality

$$|f_n - f_{n+p}| \leqq \tfrac{1}{m}$$

holds on Q for every positive integer p. Let Q_{np}^m be the set of all points of P at which the above inequality holds when m, n, p are fixed. Then Q_{np}^m is F of class α and contains Q. The product $\prod_{p=1}^{\infty} Q_{np}^m$ is likewise F of class α and must contain Q. Denote by Q_m the product of the sets

$$Q_{n_m}^m Q_{n_m+1}^m \cdots \quad Q_{n_m+1}^m.$$

The product $\overline{Q} = \prod_{m=1}^{\infty} Q_m$ is the set required by the theorem. We must show that the convergence is uniform on $\overline{Q}$. It is of course obvious that $\overline{Q}$ is F of class α and contains Q.

Let e be an arbitrary small positive number and choose m so that $\frac{1}{m} \leqq e$. Then for all points of Q_m, we have, whenever $n \geqq n_m$,

$$|f_n - f_{n+p}| \leqq e \ (p = 1, 2, 3, \ldots).$$

But $\overline{Q}$ is contained in $Q_m(m)$. Therefore this inequality holds on $\overline{Q}$ for every e. This establishes the desired uniformity.

COROLLARY I. *If a sequence of continuous functions converges uniformly on a set Q it converges uniformly on Q°, the set obtained by adding its limiting points to the set Q.*

This is the case $\alpha = 0$, and the sets which are F of class zero are closed. The set Q° is contained in every closed set which contains Q. The corollary may be proved directly from the fact that if a function f is continuous and $|f| \leqq e$ at the points of Q, then the same inequality is true for all points of Q°.

COROLLARY II. *In case $\alpha = 0$, the limit function is continuous on Q and Q° and the function on Q is the reduction relative to Q of a continuous function.*

12. THEOREM. *If a sequence $f_1, f_2, \ldots, f_n, \ldots$ of functions of class $\leqq \alpha$ converges uniformly on a set Q which is F of class α to a limit function f, the set QE, where $E = (a \leqq f)$ is F of class α.*

Because of the uniform convergence on Q, we have for every positive integer

functions of class A_α at most, and $P = \sum_{m=1}^{\infty} Q_m$, then

$$f = \sum_{m=1}^{\infty} \varphi_m f_m$$

is of class A_α on P.

Consider the set $E = (a \leqq f \leqq b)$. It is of the form $E = \sum_{m=1}^{\infty} U_m$, where $U_m = (a \leqq \varphi_m f_m \leqq b)$ and is A of class α (m). But $Q_m - U_m$ is A of class $\alpha(m)$ and therefore both $P - E = \sum_{m=1}^{\infty} (Q_m - U_m)$ and E are O of class α. It follows that E is A of class α, which was to be proved.

10. Reduction-functions of class α on a subset of P. If a function f is defined on P its reduction $f(Q)$ relative to a subset Q of P is the function on Q which has the functional values of f.*

We shall find it convenient to classify functions on an arbitrary subset Q of the fundamental set.

A function f is of class α on Q in case it is not of class $< \alpha$ on Q and is the reduction relative to Q of a function ψ which is the limit on P of a sequence of functions of class $\leqq \alpha$ on P converging uniformly on Q.

It is evident that the reduction relative to Q of a function of class α on P is of class $\leqq \alpha$ on Q, but that a function of class α on Q may not be the reduction of a function of class α, although it must be the reduction of a function of class $\alpha + 1$. It will be noticed that the reduction of a function of class α may be of lower class; for example, the reduction relative to Q of the characteristic of Q is of class zero. Not every continuous function on an open set is of class zero.† For example, let Q denote the sequence $1, \frac{1}{2}, \cdots, \frac{1}{n}, \cdots$, and let $f(\frac{1}{n})$ equal 1 (n odd), 2 (n even). Then $f(\frac{1}{n})$ is of class 1, although continuous on Q.

It follows readily from a theorem of Vallée-Poussin (loc. cit., § 130, p. 130) that the class α of functions on Q as defined above has the usual closure relative to uniform convergence. The functions of class $\alpha + 1$ whose reductions relative to Q are of class α on Q form a subclass of the class $\alpha + 1$.

The definition of function of class α on Q given above is to be compared with the following definition of Vallée-Poussin (loc. cit., § 135, p. 139): *A function f is within e of class $\alpha > 0$ on a set Q if, on Q, f is within e of being a limit of functions of class $< \alpha$ on P.*

The class of functions, within e of class α on Q for every $e > 0$, contains the functions of class α on Q as defined above. If Q is a closed set these two classes

* E. H. Moore, loc. cit., p. 91, § 52.

† It is easy to see that every function which is continuous on an arbitrary set Q is the reduction of a semi-continuous function, that is, a function of class A_1.

of functions are the same for every value of $\alpha > 0$. It would be of interest to know whether there exist classes Q for which this conclusion fails to hold.

11. THEOREM. *Let Q be an arbitrary subset of P and suppose that $f_1, f_2, \ldots, f_n \ldots$, is a sequence of functions of class $\leqq \alpha$ converging uniformly on Q. Then Q is contained in a set $\overline{Q}$ which is F of class α and the sequence converges uniformly on $\overline{Q}$.*

Because of the uniform convergence on Q we have for every integer m and integer n_m such that for all $n \geqq n_m$ the inequality

$$|f_n - f_{n+p}| \leqq \frac{1}{m}$$

holds on Q for every positive integer p. Let $Q_{n_p}^m$ be the set of all points of P at which the above inequality holds when m, n, p are fixed. Then $Q_{n_p}^m$ is F of class α and contains Q. The product $\prod_{p=1}^{\infty} Q_{n_p}^m$ is likewise F of class α and must contain Q. Denote by Q_m the product of the sets

$$Q_{n_m}^m Q_{n_m+1}^m \cdots \quad Q_{n_m+1}^m.$$

The product $\overline{Q} = \prod_{m=1}^{\infty} Q_m$ is the set required by the theorem. We must show that the convergence is uniform on $\overline{Q}$. It is of course obvious that $\overline{Q}$ is F of class α and contains Q.

Let e be an arbitrary small positive number and choose m so that $\frac{1}{m} \leqq e$. Then for all points of Q_m, we have, whenever $n \geqq n_m$,

$$|f_n - f_{n+p}| \leqq e \ (p = 1, 2, 3, \ldots).$$

But $\overline{Q}$ is contained in $Q_m(m)$. Therefore this inequality holds on $\overline{Q}$ for every e. This establishes the desired uniformity.

COROLLARY I. *If a sequence of continuous functions converges uniformly on a set Q it converges uniformly on Q°, the set obtained by adding its limiting points to the set Q.*

This is the case $\alpha = 0$, and the sets which are F of class zero are closed. The set Q° is contained in every closed set which contains Q. The corollary may be proved directly from the fact that if a function f is continuous and $|f| \leqq e$ at the points of Q, then the same inequality is true for all points of Q°.

COROLLARY II. *In case $\alpha = 0$, the limit function is continuous on Q and Q° and the function on Q is the reduction relative to Q of a continuous function.*

12. THEOREM. *If a sequence $f_1, f_2, \ldots, f_n, \ldots$ of functions of class $\leqq \alpha$ converges uniformly on a set Q which is F of class α to a limit function f, the set QE, where $E = (a \leqq f)$ is F of class α.*

Because of the uniform convergence on Q, we have for every positive integer

m a positive integer m_n such that for all points of QE, $f_{nm} \geqq a - \frac{1}{m}$. Let $E_m = (f_{nm} \geqq a - \frac{1}{m})$ and set $Q_m = QE_m$. Then Q_m is F of class α and contains QE. In fact $QE = \prod_{n=1}^{m} Q_n$, since if M is a point common to all Q_m then

$$f(M) = \lim_{M \to \infty} f_{nm}(M) \geqq a$$

and M is by definition a point of E.

13. Theorems on superposition generalizing theorems of Baire.*

THEOREM I. *If f_1 and f_2 are of class $\leqq \alpha$ and Q_1, Q_2 are A of class α the function f equal to f_1 on Q_1 and to f_2 on $Q_2 - Q_1 Q_2$ is of class $\leqq \alpha$ on $Q_1 + Q_2$.*

The characteristics φ_1, φ_2 of Q_1, $Q_2 - Q_1Q_2$ are of class $\leqq \alpha$. The function

$$f = \varphi_1 f_1 + \varphi_2 f_2$$

is of class $\leqq \alpha$ and satisfies the conditions of the theorem. The theorem may obviously be extended to any finite number of sets and functions.

THEOREM II. *If $f_0, f_1, f_2, \ldots, f_n, \ldots$, is a sequence of functions of class $\leqq \alpha$ on $P = P_0$ and if $P_1, P_2, \ldots, P_n, \ldots$ are subsets of P and A of class $< \alpha$, the function f equal to f_n on the set $P_n(P_0 - \sum_{i=1}^{n} P_i)$ and to f_0 on $P_0 - \sum_{n=1}^{\infty} P_i$ is of class $\leqq \alpha$.*

Evidently the characteristic φ_n of $P_n(P_0 - \sum_{i=1}^{n-1} P_i)$ is of class $< \alpha$. For each value of n there is a sequence of functions f_{km} such that $\lim_{k \to \infty} f_{kn} = f_n$. The function φ_ν equal to $f_{\nu n}$ on $P_n(P_0 - \sum_{i=1}^{n-1} P_i)$ for $n = 1, 2, \ldots, \nu$, and to $f_{\nu 0}$ on $P_0 - \sum_{i=1}^{\nu} P_i$ is, by an obvious extension of the preceding theorem, of class $< \alpha$. But $\lim_{\nu \to \infty} \varphi_\nu = f$, and therefore f is of class $\leqq \alpha$.

IV. CONVERGENCE (R) — LIMIT FUNCTIONS OF SEQUENCES CONVERGENT (R) — CLASSIFICATION OF FUNCTIONS IN TERMS OF CONVERGENCE (R)

14. Definition of convergence (R). Theorem of equivalence.

According to the definition of E. H. Moore a sequence of functions f_n ($n = 1, 2, 3, \ldots$) is relatively uniformly convergent, that is, convergent (R), on a range P (which may be an arbitrary set) if there exist functions f, σ defined on

* Acta Mathematica, vol. 30 (1906), pp. 16, 17, § 15; p. 31, § 28.

P, and for every small positive number e an integer n_e, such that the inequality

$$|f_n - f| \leqq e|\sigma|$$

holds for all elements of P. If the scale function σ is identically 1 the convergence becomes uniform convergence. If σ is bounded from zero and infinity the convergence is equivalent to uniform convergence. If σ is a limited function and not bounded from zero the convergence is subject to a stronger condition than that of uniform convergence. Finally if σ is unlimited and $|\sigma| > a > 0$ the condition is weaker than the condition of uniform convergence. It is possible to define convergent sequences which do not converge uniformly relative to any scale function σ.*

The following relation between uniform and relatively uniform convergence is of fundamental importance.

THEOREM. *A necessary and sufficient condition that a convergent sequence of functions f_n converge relatively uniformly on P to a function f is that there exist a sequence of classes Q_m such that Q_{m+1} contains Q_m, every element of P lies in some class Q_m and the sequence converges uniformly on Q_m (m).*†

The condition is necessary. From the existence of the scale function σ we obtain Q_m as the set $(|\sigma| \leqq m)$.

The condition is sufficient. Let φ denote the function whose value is equal to the greatest of the values of the functions in the convergent sequence, $|f_1|, |f_2|, \ldots, |f_n| \ldots$. The function φ is everywhere defined and finite, and is in fact the smallest function for which all of the inequalities

$$|f_n| \leqq \varphi \ (n = 1, 2, 3, \ldots)$$

hold uniformly on P.

15. Theorems on convergence (R).

THEOREM I. *If* $f = \lim_{n\to\infty} f_{1n}$ *and* $f_2 = \lim_{n\to\infty} f_{2n}$ *and the convergence is relatively uniform in each case then* $f_1 + f_2 = \lim_{n\to\infty} f_{1n} + f_{2n}$ *and the convergence is relatively uniform.*

In fact if the scale functions are, respectively, σ_1, σ_2, then $\sigma = |\sigma_1| + |\sigma_2|$ is effective as a scale function in the case of the sequence of functions, $f_{1n} + f_{2n}$.

It is evident that this theorem may be extended to the case of any finite number of functions.

---

* These Transactions, vol. 15, p. 201.

† "Relatively uniform convergence" and "relatively uniform convergence with respect to σ" are different concepts. In one case there exists a scale function σ, in other the scale function is prescribed.

E. H. Moore has called attention to the following. *If* $f_m = \lim_{n\to\infty} f_{mn}$ *and the convergence is uniform relative to* $\sigma_m (m = 1, 2, \ldots)$, *and if there is a function* σ *and a sequence of positive numbers* a_m *such that* $|\sigma_m| \leqq a_m|\sigma|$ *identically for every* m, *then the convergence to* f_m *is in every case uniform relative to* σ.*

The function σ is said to be a dominant of the sequence of functions σ_m.

THEOREM II. *If* $f = \lim_{m\to\infty} f_m$ *uniformly relative to* σ_0, *and* $f_m = \lim_{n\to\infty} f_{mn}$ *uniformly relative to* $\sigma_m (m = 1, 2, \ldots,)$ *and if* σ *is a dominant of the sequence* $\sigma_0, \sigma_1, \ldots, \sigma_m, \ldots$, *then a subsequence of the sequence of functions* f_{mn} *converges to* f *uniformly relative to* σ.

From the theorem of Moore just stated each convergence is uniform relative to σ. Choose f_{mn_m} so that $|f_m - f_{mn_m}| \leqq \frac{1}{m}|\sigma|$. Then the functions f_{mn_m} converge to f uniformly relative to σ.

THEOREM III. *If* $E_1, E_2, E_3, \ldots$ *are non overlapping sets, and* $f_1, f_2, f_3, \ldots$ *is any sequence of functions, the series*

$$\sum_{m=1}^{\infty} f_m \varphi_m,$$

where φ_m *is the characteristic of* E_m, *converges relatively uniformly on* P.

Let $E = \sum_{m=1}^{\infty} E_m$. Let $Q_m = \sum_{i=1}^{m} E_i + P - E$. Then $\lim_{n\to\infty} Q_m = P$ and the convergence is uniform on Q_m (m).

THEOREM IV. *If a sequence* $f_1, f_2, f_3, \ldots$ *of functions converges* (R) *to a function* f *and if* φ *is any uniformly continuous function over the interval of variation of the functions* $f, f_1, f_2, \ldots$, *then* $\varphi(f) = \lim_{n\to\infty} \varphi(f_n)$ *and the convergence is relatively uniform.*

For f_n converges to f uniformly on each of a sequence of sets $Q_1, Q_2, Q_3, \ldots, Q_m, \ldots$ such that Q_m contains Q_{m-1} and $\lim_{n\to\infty} Q_m = P$. It follows at once that $\varphi(f_n)$ converges uniformly to $\varphi(f)$ on Q_m. Therefore the convergence is relatively uniform.

16. On the limits of relatively uniformly convergent sequences of functions of class $\leqq \alpha$.

THEOREM I. *If* f *is the limit of a sequence of functions* f_n *of class* $\leqq \alpha$ *converging* (R) *there exists a sequence* $Q_1, Q_2, Q_3, \ldots$, *of subsets of* P *such that:* Q_{m+1} *contains* Q_m, Q_m *is* F *of class* α; *the convergence is uniform on* Q_m; f *is of class* $\leqq \alpha$ *on* $Q_m(m)$; *and* $\lim_{m\to\infty} Q_m = P$.

This theorem is an immediate consequence of the theorems of §12 and §14.

* E. H. Moore, loc. cit., p. 50, § 25, Theorem I.

COROLLARY I. *Every sequence of functions of class* $\leqq \alpha$, *which converges* (R), *converges uniformly relative to a scale function of the class* $A_{\alpha+1}$ *at most.*

COROLLARY II. *If the sets* Q_m *are* A *of class* α *there is an effective scale function of class* D_α *at most.*

THEOREM II. *A necessary and sufficient condition that a function* f *be the limit of a sequence of functions of class* $\leqq \alpha$ *convergent* (R) *is that there exist a sequence*

$$Q_1, Q_2, Q_3, \ldots,$$

of subclasses of P *such that* $\sum_{m=1}^{\infty} Q_m = P$, *and the reduced function* $f(Q_m)$ *shall be of class* $\leqq \alpha$ (m).

The necessity of the condition follows from the preceding theorem. To prove the condition sufficient we note that there is for every positive integer m a function f_m of class $\leqq \alpha$ such that on $\sum_{i=1}^{m} Q_i$, $|f - f_m| \leqq \frac{1}{m}$. The sequence of functions f_m converges uniformly on $\sum_{i=1}^{n} Q_i$ for every value of n and is therefore convergent (R) by the theorem of § 14.

THEOREM III. *If a sequence of functions* f_n *of class* $\leqq \alpha$ *is convergent* (R) *to* f *every set* $E = (a \leqq f)$ *is the limit of sets* F *of class* α. *Furthermore, there exists (for every value of* a*) a sequence of sets* Q_m *which are* F *of class* α *and have the sum* P, *such that* EQ_m *is* F *of class* α *for every value of* m $(m = 1, 2, 3, \ldots)$.

The corresponding statements may be made for the sets $E = (f \leqq a)$.

From the theorem of § 16 there exists a sequence of sets Q_m with limit P such that each set Q_m is F of class α and the convergence is uniform on Q_m. From the theorem of § 13 the sets Q_mE are F of class α. Since $\lim_{m\to\infty} Q_mE = E$, the theorem is proved.

COROLLARY. *The limit of a sequence of functions of class* $\leqq \alpha$ *convergent* (R) *is of class* $A_{\alpha+1}$ *at most.*

For the sets EQ_m are F of class α and therefore O of class $\alpha+1$. But $E = \sum_{m=1}^{\infty} EQ_m$ and is therefore O of class $\alpha + 1$. Since the limit function is of class $\alpha+1$ at most, the sets E are F of class $\alpha+1$, therefore A of class $\alpha+1$, which implies the conclusion stated.

17. Theorems on the limit of a sequence of functions of class $< \alpha$ convergent (R).

In case α is of the second kind we obtain a more precise result than is obtained from the statement of the theorem of § 16.

THEOREM I. *If a function f is the limit of a sequence of functions f_n of class $< \alpha$ convergent (R), every set $E = (a \leqq f)$ is A of class α.*

Since the convergence is uniform on each of a sequence of sets Q_m we may find n_{km} such that on Q_m,

$$|f_n - f_{n+k}| \leqq \tfrac{1}{k} \quad (k = 1, 2, 3, \ldots),$$

whenever n exceeds n_{km}.

Let Q_{nh}^{km} denote the set of all points of P at which the preceding inequality holds. Evidently Q_{nh}^{km} is A of some class $\beta < \alpha$. From the quadruply infinite system $((Q_{nh}^{km}))$ we may select a subsystem (T^{km}) such that the product T^m of the T^{km} contains Q_m while the sequence f_n converges uniformly on T^m. This is an obvious consequence of the fact that the inequality above holds on T^{km} which must contain Q_m. If we set $E_{nk} = (a - \frac{1}{k} \leqq f_n)$ then E_{nk} is A of some class $\beta < \alpha$. Therefore $T^{km}E_{nk}$ is A of some class $\beta < \alpha$. But $T^mE = \prod_{k=1}^{\infty} T^{km}E_{n_k k}$ where n_k is chosen so that E_{n_k} contains Q_mE, which choice can be made because of the uniform convergence on Q_m. It follows that T^mE is A of class α and so O of class α. But $E = \sum_{m=1}^{\infty} T^mE$ is of the form $\sum O_\alpha$, that is, E is O of class α. But E is also F of class α, since f is of class α and $E = (a \leqq f)$. Therefore E is A of class α.

From the preceding theorems and the corollary to a theorem of § 16 we obtain the important result:

THEOREM II. *If a sequence of functions f_n of class $< \alpha$ converges to a limit function f of class α relatively uniformly, then f belongs to the class A_α.*

18. Sequences of characteristic functions, convergent (R).

THEOREM I. *If $Q = \lim_{m \to \infty} Q_m$ then the characteristics φ_m of the sets Q_m converge relatively uniformly to the characteristic of Q.*

In fact on $QQ_m + (P - Q)(P - Q_m)$ we have $\varphi = \varphi_m$, where φ denotes the characteristic of Q. Let $V_m = \prod_{n=m}^{\infty} [QQ_m + (P - Q)(P - Q_m)]$. It is at once evident that the convergence is uniform on $V_m(m)$, and that V_m contains V_{m-1}. We will show that every element of P lies in V_m for a sufficiently large value of m. In fact if M is a point of Q there is, since $Q = \lim_{m \to \infty} Q_m$, an integer m_0 such that for all $m \geqq m_0$, M belongs to Q_m and therefore to QQ_m, and so to V_{m_0}. If M is a point of $P - Q$ a similar argument prevails.

THEOREM II. *Every function of class D_α is a limit of a relatively uniformly convergent sequence of functions of class $< \alpha$.*

Suppose f of class D_α is equal to a_m on E_m, ambiguous of class α. Let φ_m be the characteristic of E_m and consider the function

$$g_m = \sum_{k=1}^{m} a_k \varphi_k.$$

Since E_m is the limit of a sequence of sets A of class $< \alpha$ it follows that φ_k $(k \leqq m)$ is the limit of a sequence φ_{kn} of characteristic functions of class $< \alpha$ convergent (R). Therefore g_m is the limit as n becomes infinite of the sequence $g_{mn} = \sum_{k=1}^{m} a_k \varphi_{kn}$, convergent (R). We proceed to show that a sequence taken from the doubly infinite set of functions g_{mn} converges (R) to f.

Since $E_m = \lim_{n \to \infty} E_{mn}$, where E_{mn} is A of class $< \alpha$, the functions φ_{mn} are equal to 1 and to φ_m on $H_{mk} = \coprod_{i=k}^{\infty} E_{mi}$ whenever n exceeds k. Evidently $E_m = \lim_{k \to \infty} H_{mk}$. It follows that

$$f = g_m = g_{mn}$$

on $Q_m = \sum_{k=1}^{m} H_{km}$. But $\lim_{m \to \infty} Q_m = P$. For if M is a point of P it lies in some class E_{m_0}, therefore in some set $H_{m_0 k_0}$. Then if $m_0 \geqq k_0$, the point M belongs to Q_{m_0}. If $m_0 < k_0$, the point M belong to Q_{k_0}. Therefore the sequence g_{mm} converges (R) to the function f.

19. The classification of function in terms of convergence (R).

We have seen that the limit functions of sequences of functions of class $\leqq \alpha$ which are convergent (R) belong to class $A_{\alpha+1}$ at most. Denoting the class of all functions of class α by B_α we see that while B_α is not closed for relatively uniform convergence its extension in terms of relatively uniform convergence is a subclass of $B_\alpha + A_{\alpha+1}$.

If we set $R_0 = B_0$; R_1 equal to the class of all functions which are not in R_0 and are the limits of relatively uniformly convergent sequences of functions of R_0; and define R_α in similar fashion in terms of the ordinals $< \alpha$; then it is evident that

$$D_{\alpha+1} < R_{\alpha+1} \leqq B_\alpha + A_{\alpha+1}.$$

In case α is of the second kind we have

$$D_\alpha < R_\alpha \leqq A_\alpha < B_\alpha.$$

We have established previously* that

$$R_1 = A_1.$$

* These Transactions, vol. 20 (1919), p. 184.

UNIVERSITY OF IOWA,
IOWA CITY, IOWA

PERIODIC FUNCTIONS WITH A MULTIPLICATION THEOREM *

BY

J. F. RITT

I. Introduction

Let $f(z)$ represent a uniform analytic function with no essential singularity in the finite part of the plane, and periodic, either simply or doubly. There exists in certain cases a number m, different from unity, such that $f(mz)$ is a rational function of $f(z)$. It is desired to determine all of those cases.

Thus a part of our problem is the determination of all elliptic functions which admit of complex multiplication. As far as the writer knows, this question has not been investigated before.

It will be seen that we must have $|m| \geqq 1$, and that the cases with $|m| = 1$ are easy to handle. The problem obtains its interest when $|m|$ exceeds unity, and then it can be stated in a form which shows its relation to a result of Poincaré.† He has shown that if $R(z)$ is a rational function such that $R(0) = 0$, and that $R'(0)$, which we shall denote by m, has a modulus greater than unity, there exists a function $f(z)$, meromorphic or entire, such that

$$f(mz) = R[f(z)].$$

It would be a problem equivalent to the one stated to determine the cases in which $f(z)$ is periodic.

The results are almost negative. We shall state them in detail for $|m| > 1$.

If $f(z)$ is simply periodic, $f(z)$ is a linear function, either of $\cos(\alpha z + \beta)$, where β is restricted to certain values, or of $e^{\alpha z}$. It will perhaps be considered a noteworthy fact that $f(z)$ cannot be a transcendental function of an exponential, as are, for example, the theta functions.

If $f(z)$ is doubly periodic, and not of the lemniscatic or equianharmonic types, it is of the form $\wp(z + \alpha)$, where α is restricted to certain values. These are the functions regularly used in the theory of complex multiplication.

For the lemniscatic case there are functions of the form $\wp^2(z + \alpha)$ in addition to those of the form $\wp(z + \alpha)$, while for the equianharmonic case the forms

* Presented to the Society, February 26, 1921.

† H. Poincaré, *Sur une classe nouvelle de transcendantes uniformes*, Journal de Mathématiques, ser. 3, vol. 55 (1890).

$\wp^3(z+\alpha)$ and $\wp'(z+\alpha)$ are possible. These special cases have been noticed before.*

The possibilities for $|m| = 1$ are enumerated rapidly in §VII.

It will be noticed that we actually use, with respect to $f(mz)$, only the condition that it be a uniform function of $f(z)$.

II. Proof that $|m| \geqq 1$

It is almost obvious that the multiplier m must not be less than unity in modulus. Suppose, in fact, that $|m| < 1$. If the value $f(0)$ is assumed n times by $f(z)$ at the point $z = 0$, we can determine a neighborhood of the origin in which no value of $f(z)$ is assumed more than n times. Now if h is a period of $f(z)$, and z_0 any value of z, we have

$$f(z_0) = f(z_0 + h) = \cdots = f(z_0 + ih) = \cdots,$$

and, since $f(mz)$ is a uniform function of $f(z)$,

$$f(m^p z_0) = f[m^p(z_0 + h)] = \cdots = f[m^p(z_0 + ih)] = \cdots,$$

where p is any integer. If we had $|m| < 1$, we could, by taking p sufficiently large, bring an unlimited number of points $m^p(z_0 + ih)$ into the neighborhood taken above, which is impossible, since $f(z)$ is the same for all of those points.

III. The auxiliary automorphism

Assuming that $|m|$ exceeds unity, which, in fact, is generally the case in multiplication theorems, we shall deduce a relation which will permit us to show that $f(z)$ cannot take on any value more than twice in a period strip if $f(z)$ is simply periodic, and, except in two cases, not more than twice in a period parallelogram if $f(z)$ is doubly periodic.†

Of course if $f(z)$ is doubly periodic, it assumes all values the same number of times in a period parallelogram. If $f(z)$ is simply periodic, and is a rational function of an exponential, it assumes all values the same number of times in a period strip, it being understood that two infinite points are adjoined to the strip. If $f(z)$ is simply periodic and is a transcendental function of an exponential, we know from Picard's theorem that $f(z)$ must assume all values, except two, possibly, an infinite number of times in each period strip. These details being mentioned, we shall, in what follows immediately, use the term "fundamental domain" to designate a primitive period strip or a primitive period parallelogram according as $f(z)$ is simply periodic or doubly periodic.

---

* Weber, *Algebra*, vol. 3, p. 572.

† It is understood that we refer to *primitive* periods.

Taking a fundamental domain which contains the origin in its interior, suppose that in it $f(z)$ assumes values more than once.

Let h be *any* period of $f(z)$. Consider the points

$$0, \frac{h}{m}, \frac{h}{m^2}, \cdots, \frac{h}{m^p}, \cdots,$$

which accumulate at the origin. On the points of this infinite sequence, $f(z)$ must assume an infinite number of distinct values; also, $f'(z)$, the derivative of $f(z)$, can vanish for only a finite number of these points. There must consequently exist a point z_1 of the above sequence, interior to the fundamental domain, at which $f(z)$ is analytic, at which $f'(z)$ is not zero, and at which $f(z)$ does not assume an exceptional value, if exceptional values exist; that is, there is another point z_2 of the fundamental domain for which $f(z)$ has the same value as at z_1. If $f'(0) \neq 0$ and if $f(0)$ is neither infinity nor an exceptional value of $f(z)$, it is simplest to take $z_1 = 0$.

Consider any neighborhood Γ_1 of z_1 in which $f(z)$ assumes no value more than once; the existence of such a neighborhood follows from the fact that $f'(z_1)$ is not zero. There is a neighborhood Γ_2 of z_2 in which $f(z)$ assumes no value which it does not take on in Γ_1.

Associate with every point of Γ_2 the affix of that point of Γ_1 at which $f(z)$ assumes the same value. This defines a function $\varphi(z)$, analytic in Γ_2. It has a Taylor development

$$\varphi(z) = z_1 + \alpha_1(z - z_2) + \alpha_2(z - z_2)^2 + \cdots + \alpha_n(z - z_2)^n + \cdots \tag{1}$$

We are going to show that $\varphi(z)$ is actually linear.

Let r be an integer so great that of the three points

$$z_2 + \frac{h}{m^r}, \qquad z_1 + \frac{h}{m^r}, \qquad \varphi\left(z_2 + \frac{h}{m^r}\right),$$

the first lies in Γ_2 and the second in Γ_1. Also if $z_1 \neq 0$, we suppose that r exceeds the integer p used in defining z_1.

We have

$$f\left[\varphi\left(z_2 + \frac{h}{m^r}\right)\right] = f\left(z_2 + \frac{h}{m^r}\right). \tag{2}$$

Understanding that $f(mz) = R[f(z)]$, let $R_n(z)$ denote the nth iterate of $R(z)$. We have

$$f(m^n z) = R_n[f(z)]. \tag{3}$$

It is clear then, from (2) and (3), that

$$f\left[m^r\varphi\left(z_2+\frac{h}{m^r}\right)\right]=f\left[m^r\left(z_2+\frac{h}{m^r}\right)\right]. \tag{4}$$

But

$$f\left[m^r\left(z_2+\frac{h}{m^r}\right)\right]=f(m^rz_2+h)=f(m^rz_2)$$
$$=R_r[f(z_2)]=R_r[f(z_1)].$$

We have thus, by (4),

$$f\left[m^r\varphi\left(z_2+\frac{h}{m^r}\right)\right]=R_r[f(z_1)]=f(m^rz_1). \tag{5}$$

It is desirable now to show that mh is a period of $f(z)$. This is obvious since

$$f(z+mh)=R\left[f\left(\frac{z}{m}+h\right)\right]=R\left[f\left(\frac{z}{m}\right)\right]=f(z).$$

Similarly m^rh is a period, r being any positive integer.

As r exceeds p, we see that

$$f(m^rz_1)=f(m^{r-p}h)=f(0).$$

We have by (1)

$$m^r\varphi\left(z_2+\frac{h}{m^r}\right)=m^{r-p}h+\alpha_1h+\alpha_2\frac{h^2}{m^r}+\cdots, \tag{6}$$

and by (5) and (6), remembering that $m^{r-p}h$ is a period,

$$f\left(\alpha_1h+\alpha_2\frac{h^2}{m^r}+\cdots\right)=f(0), \tag{7}$$

for any r greater than p.

It follows from (7) that $\alpha_n=0$ for $n>1$, otherwise, allowing r to pass through all values greater than p, we would find an infinite number of distinct points, accumulating at the point α_1h, for all of which $f(z)$ assumes the value $f(0)$, and α_1h would be an essential singularity of $f(z)$. Evidently α_1 cannot be zero.

Thus a relation

$$f[z_1+\alpha_1(z-z_2)]=f(z) \tag{8}$$

holds for the region Γ_2. Since the functions which appear in it are analytic, it must hold for the whole plane.

Replacing z by z/α_1+z_2 and $1/\alpha_1$ by ϵ, we may write (8) in the simpler form

$$f(z_1+z)=f(z_2+\epsilon z). \tag{9}$$

This is the relation referred to at the beginning of the present section.

IV. First consequences of the auxiliary automorphism

It follows from (9) that if h is any period of $f(z)$ (not necessarily the one used in deriving (9)), ϵh and h/ϵ are also periods. Replacing z by $z + h$ in (9), we have

$$f(z_1 + z + h) = f(z_2 + \epsilon z + \epsilon h), \tag{10}$$

and remembering that since h is a period, the first member of (10) equals the second of (9), we see that

$$f(z_2 + \epsilon z) = f(z_2 + \epsilon z + \epsilon h),$$

from which it follows that ϵh is a period. Similarly, replacing z by $z + h/\epsilon$ in (9), we find

$$f\left(z_1 + z + \frac{h}{\epsilon}\right) = f(z_2 + \epsilon z + h) = f(z_1 + z),$$

so that h/ϵ is a period.

Thus $\epsilon^r h$, r being any positive or negative integer, is a period of $f(z)$. It follows at once that $|\epsilon| = 1$, for if not, the moduli of the periods $\epsilon^r h$ could be made small at pleasure. Furthermore, ϵ must be a root of unity, otherwise the affixes of the periods $\epsilon^r h$ would be dense along a circle. Finally we cannot have $\epsilon = 1$; indeed, the fundamental relation (9) shows that in that case $z_2 - z_1$ would be a period, which is certainly impossible, since z_1 and z_2 are both interior to a primitive period strip or to a primitive parallelogram.

These facts in hand, it will not be hard to show that if $f(z)$ is simply periodic ϵ equals -1, and that if $f(z)$ is doubly periodic, ϵ, if not -1, is either a third, fourth or sixth root of unity.

The truth of the statement for the simply periodic case follows from the fact that if h is a period of smallest modulus, ϵh, being also a period, must be $-h$.

For the doubly periodic case the proof depends on the proper choice of a primitive pair of periods. We proceed as follows. Of all the periods of $f(z)$, there are two or more which have a smallest modulus. Let one of these be taken as a first period $2\omega_1$. Rejecting all real multiples of $2\omega_1$, there are two or more of those which remain which have a smallest modulus. Of these we choose one, calling it $2\omega_3$. Then $2\omega_1$ and $2\omega_3$ are a primitive pair of periods. For let a parallelogram be constructed with one vertex at the origin and with sides $2\omega_1$ and $2\omega_3$. If these periods were not primitive, the parallelogram would contain the affixes of other periods, of which at least one would have a modulus less than $|2\omega_1|$ or less than $|2\omega_3|$.

Consider the case where ϵ is not -1. One of the four numbers

$$\epsilon, \quad -\epsilon, \quad 1/\epsilon, \quad -1/\epsilon,$$

has a positive argument not greater than $\pi/2$. Let that number be denoted by α and its argument by θ. Having chosen $2\omega_1$ we can take $2\alpha\omega_1$ as the second period.

If $\theta = \pi/2$ the period parallelogram is a square and we have the *lemniscatic* case.

If $\theta < \pi/2$, the periods $2\omega_1$ and $2\alpha^2\omega_1$ will form a primitive pair. Since any two primitive period parallelograms have the same area, the angle 2θ of the parallelogram formed with $2\omega_1$ and $2\alpha^2\omega_1$ must be supplementary to the angle θ in the parallelogram formed with $2\omega_1$ and $2\alpha\omega_1$. We have thus $\theta = \pi/3$, which shows that we are in the *equianharmonic* case.

V. The non-exceptional cases with $|m|>1$

If $f(z)$ is not an elliptic function of the lemniscatic or equianharmonic types, we must have $\epsilon = -1$. It follows that $f(z)$ cannot take on a value more than twice in a fundamental domain, for if $f(z_1) = f(z_2) = f(z_3)$, z_1 being taken as a period divided by a power of m, and z_2 and z_3 being any two other points in the same fundamental domain with z_1, we must have, by (9),

$$f(z_1 + z) = f(z_2 - z) = f(z_3 - z),$$

which leads to the absurdity that $z_3 - z_2$ is a period.

Here are two simple observations which will make easy the enumeration of the functions $f(z)$.

If $f(mz) = R[f(z)]$, we have also, letting $F(z) = f(az)$ where a is any constant,

$$F(mz) = R[F(z)].$$

Thus, for the simply periodic case, we may suppose that the period is 2π; we will secure complete generality by multiplying the argument in $f(z)$ by an arbitrary constant.

Also, if $\varphi(z)$ is any linear function

$$\varphi(z) = \frac{az + b}{cz + d},$$

and $\varphi^{-1}(z)$ its inverse, we have

$$\varphi^{-1}f(mz) = \varphi^{-1}R\varphi[\varphi^{-1}f(z)],$$

so that from any function with a multiplication theorem, we can derive others, perhaps simpler ones, which also have multiplication theorems.

Suppose then that $f(z)$ takes on all values twice in a fundamental domain. According to (9) we have

$$f(z_1 + z) = f(z_2 - z),$$

z_1 and z_2 being properly chosen. Replacing z by $z - (z_1 - z_2)/2$ and $(z_1 + z_2)/2$ by u, we have

$$f(u + z) = f(u - z). \tag{11}$$

Consider the case where $f(z)$ is simply periodic with the primitive period 2π. We see from (11) that $f(u + z)$, being even with respect to z, and taking on all values twice in a strip, is a linear function, integral or fractional, of $\cos z$. What is the same, $f(z)$ is a linear function of $\cos(z - u)$. Then some linear function of $f(z)$, which also has a multiplication theorem, equals $\cos(z - u)$. It remains to determine when $\cos(mz - u)$ is a rational function of $\cos(z - u)$. Of course m must be an integer, since $2m\pi$ has to be a period of $\cos(z - u)$. Furthermore, since $\cos(z_1 - u) = \cos(z_2 - u)$ when

$$z_1 - u + z_2 - u \equiv 0 \pmod{2\pi}, \tag{12}$$

we must have, in that case,

$$mz_1 - u + mz_2 - u \equiv 0 \pmod{2\pi}. \tag{13}$$

Multiplying (12) by m and subtracting the result from (13), we find

$$u = \frac{k\pi}{m - 1},$$

where k is an integer.

If $f(z)$ assumes no value more than once in a period strip, it must be a linear function of e^{iz}.

Summing up, we see that the only simply periodic functions with a multiplication theorem are the functions

$$\frac{a\cos(\alpha z + \beta) + b}{c\cos(\alpha z + \beta) + d}, \qquad \beta = \frac{k\pi}{m - 1},$$

and the functions

$$\frac{ae^{\alpha z} + b}{ce^{\alpha z} + d}.$$

The nature of the constants needs no explanation.

Take now the case in which $f(z)$ is doubly periodic, and of the second order.* If the periods of $f(z)$ are $2\omega_1$ and $2\omega_3$, we see by (11) that $f(u + z)$, being even and of the second order, is a linear function of $\wp(z|\omega_1, \omega_3)$. That is, some linear function of $f(z)$ is equal to $\wp(z - u)$. We get here nothing more than the ordinary

* We shall see in § VI that if $f(z)$ is of the second order, the relation (11) holds, even if $f(z)$ is lemniscatic or equianharmonic.

formulas for the multiplication of the argument, in $\wp(z+\alpha)$, either by a real or by a complex factor. We have

$$f(z) = \frac{a\wp(z+\alpha)+b}{c\wp(z+\alpha)+d},$$

the multiplier m satisfying the congruences

$$2m\omega_1 \equiv 0,\ 2m\omega_3 \equiv 0 \qquad (\text{modd } 2\omega_1, 2\omega_3),$$

and the constant α being given by the equation

$$\alpha = \frac{h\omega_1 + k\omega_3}{m-1},$$

where h and k are any integers.

VI. The special cases with $|m| > 1$

Suppose that in the relation (9),

$$f(z_1+z) = f(z_2+\epsilon z),$$

ϵ is a primitive nth root of unity. Then the order of $f(z)$ is at least n.* For replacing z by $z + (z_1 - z_2)/(\epsilon - 1)$ and $(\epsilon z_1 - z_2)/(\epsilon - 1)$ by u, we find

$$f(u+z) = f(u+\epsilon z). \tag{14}$$

This equation shows that in the development of $f(u+z)$ in powers of z at the origin, the exponents are all multiples of n. Consequently the value $f(u)$ is assumed at least n times by $f(z)$ at the point u.

If $f(z)$ is of the lemniscatic type, and of order greater than two, it is necessarily of order four. Indeed, if $f(z)$ is of order n, there exist relations

$$f(z_1+z) = f(z_2+\epsilon_2 z) = \cdots = f(z_n+\epsilon_n z),$$

where $\epsilon_2, \ldots, \epsilon_n$ are fourth roots of unity. We cannot have $\epsilon_i = \epsilon_j$, else $z_i - z_j$ would be a period. Hence n cannot exceed 4. But since only one ϵ can equal -1, one of them must be $\pm i$. Hence n is at least four.

Similarly, if $f(z)$ is of the equianharmonic type, and of order greater than 2, it must be of order 6 or of order 3 according as there is or is not an ϵ which is a primitive sixth root of unity.

---

* Cf. the footnote on p. 22.

Suppose that $f(z)$ is lemniscatic, and of order 4. There must exist a relation

$$f(z_1 + z) = f(z_2 + iz),$$

or, by (14),

$$f(u + z) = f(u + iz).$$

As was seen above, $f(u + z)$ assumes the value $f(u)$ four times for $z = 0$, and the exponents in its expansion in powers of z at the origin are all divisible by 4. Consequently either $f(u + z)$ or $1/[f(u + z) - f(u)]$ has a pole of order 4 at the origin, with a principal part of the form c/z^4. In all, we see that some linear function $F(z)$ of $f(u + z)$ has a pole of order 4 at the origin, with $1/z^4$ for principal part. But the function $\wp^2 z$ is precisely of this type.* Hence $F(z) - \wp^2(z)$, having no poles, is a constant. It follows that $f(z)$ is a linear function of $\wp^2(z-u)$.

It remains to determine, for the lemniscatic case, when $\wp^2(mz + \alpha)$ is a rational function of $\wp^2(z + \alpha)$. We must have, first, $m = p + qi$ with p and q integral, and since $\wp^2 iu = \wp^2 u$ in this case, the congruence

$$z_1 + \alpha \equiv i(z_2 + \alpha) \qquad (\text{modd } 2\omega_1, 2\omega_3),$$

must imply the congruence

$$mz_1 + \alpha \equiv i(mz_2 + \alpha) \qquad (\text{modd } 2\omega_1, 2\omega_3),$$

so that

$$\alpha = \frac{2h\omega_1 + 2k\omega_3}{(m-1)(1-i)},$$

where h and k are any integers.

The equianharmonic cases can be handled in the same way. It will suffice to state the results.

If $f(z)$ is of the third order, it must be a linear function of $\wp'(z + \alpha)$. Here $m = p + qe^{2\pi i/3}$ with p and q integral, and

$$(m - 1)(1 - e^{2\pi i/3})\ \alpha \equiv 0 \quad (\text{modd } 2\omega_1, 2\omega_3).$$

If $f(z)$ is of the sixth order it must be a linear function of $\wp^3(z + \alpha)$, or, what amounts to the same, of the fourth derivative of $\wp(z + \alpha)$, where

$$(m - 1)(1 - e^{\pi i/3})\ \alpha \equiv 0 \qquad (\text{modd } 2\omega_1, 2\omega_3)$$

* We might also have used the second derivative of $\wp z$.

VII. The case of $|m| = 1$

In the case of $|m| = 1$, since if h is a period of $f(z)$, mh is one also, it follows as in §IV that m is either -1 or a third, fourth or sixth root of unity. Let m be a primitive nth root of unity and let

$$f(mz) = R[f(z)].$$

Since

$$f(m^n z) = f(z) = R_n[f(z)],$$

the rational function $R(z)$ must be a periodic linear fraction of which the period is n, or a divisor of n. Subjecting $f(z)$ to a suitable linear transformation, we may suppose that we are in the case where

$$f(mz) = \epsilon^r f(z),$$

where ϵ is a primitive nth root of unity.

Suppose first that $f(z)$ is simply periodic; we must have $m = -1$. If $\epsilon^r = 1$, $f(z)$ is a rational function of $\cos \alpha z$, with α arbitrary. If $\epsilon^r = -1$, $f(z)$ is the product of $\sin \alpha z$ by a rational function of $\cos \alpha z$.

If $f(z)$ is elliptic, and if $m = -1$, $f(z)$ is either a rational function of $\wp z$, or the product of $\wp' z$ by such a rational function.

It is understood that in the cases we are enumerating, any linear function of (z) also has a multiplication theorem.

If $f(z)$ is lemniscatic and if m is not -1, we may suppose that $m = i$. We must have

$$f(iz) = i^r f(z).$$

If $r = 0$, $f(z)$ is a rational function of $\wp^2 z$. If $r = 1$, $f(z)$ is the product of $\wp' z$ by a rational function of $\wp^2 z$. For $r = 2$, we multiply the rational function by $\wp z$, and for $r = 3$ we multiply it by the third derivative of $\wp z$.

The equianharmonic cases are handled the same way. If m is a third root of unity, $f(z)$ is either a rational function of $\wp' z$, or the product of such a rational function by $\wp z$ or one of its derivatives. If m is a primitive sixth root of unity one would use $\wp^3 z$ instead of $\wp' z$.

Columbia University,
New York, N. Y.

NOTE ON DIRICHLET AND FACTORIAL SERIES*

BY

TOMLINSON FORT

Landau† has proved various theorems bearing on the regions of convergence of Dirichlet and factorial series. In the present note I set up a class of series including factorial series and ordinary Dirichlet series, and forming a continuous transition from one to the other. Theorems are proved which include some of Landau's theorems as very special cases.

1. Consider the series

$$\sum_{n=1}^{\infty} a_n A_n^{(k)}(z), \tag{1}$$

where

$$A_n^{(k)}(z) = \frac{k}{z+k-1} \cdot \frac{\Gamma(nk)}{\Gamma(z+nk)} \cdot \frac{\Gamma(z+k)}{\Gamma(k)}$$

whenever z, k and n have such values that this formula defines a definite number. Whenever, for a particular point (z_0, k_0, n_0), $A_n^{(k)}(z)$ is not defined by the formula but approaches a limit as (z, k, n) approaches (z_0, k_0, n_0), $A_n^{(k)}(z)$ is given at this point the limiting value.

2. When $k = 1$, (1) reduces to the usual factorial series

$$\sum_{n=1}^{\infty} a_n \frac{(n-1)!}{z(z+1)\cdots(z+n-1)}. \tag{2}$$

3. When $k \longrightarrow \infty$, so that $|\text{am } k| \leqq \pi - \epsilon_1$, $\epsilon_1 > 0$, (1) reduces to the ordinary Dirichlet series

$$\sum_{n=1}^{\infty} \frac{a_n}{n^z}. \tag{3}$$

To prove this statement consider $A_n^{(k)}(z)$ for values of k not equal to zero and so large in absolute value that $z + nk - 1 \neq 0$ $(n = 1, 2, 3, \ldots)$ and

* Presented to the Society, December 31, 1919.

† Sitzungsberichte der mathematisch-physikalischen Klasse der Königlichen Bayerischen Akademie der Wissenschaften zu München, vol. 36 (1906), pp. 151–218; see especially pp. 167–184.

$|\text{am}(z + nk - 1)| \leqq \pi - \epsilon_2$, $\epsilon_2 > 0$. Then

$$A_n^{(k)}(z)n^z = \frac{k}{z + k - 1} \cdot \frac{\Gamma(nk)(nk)^z}{\Gamma(z + nk)} \cdot \frac{\Gamma(z + k)}{\Gamma(k)k^z}. \tag{4}$$

Consider the asymptotic form,*

$$\Gamma(z) = e^{(z-\frac{1}{2})\log z - z + \log\sqrt{2\pi} + \omega(z)},$$

where $\omega(z) = C_1/z + f_1(z)/z^2$, C_1 being a constant and $|f_1(z)| < C_2$, a constant, which asymptotic form is valid when $|\text{am } z| < \pi - \epsilon_3$, $\epsilon_3 > 0$. From it one can readily establish the following formula which is fundamental in this paper,

$$\frac{q^z\Gamma(q)}{\Gamma(q + z)} = 1 + \frac{z - z^2}{2q} + \frac{\psi(z, q)}{q^2}, \tag{5}$$

where $|\psi| < M$, a constant, which form is valid over a region where $q \neq 0$, $q + z \neq 0$ and where $|\text{am } q|$ and $|\text{am}(q + z)|$ are both less than π minus some positive constant. From (4) and (5) follows the desired result that when $k \longrightarrow \infty$

$$A_n^{(k)}(z)n^z \longrightarrow 1. \tag{6}$$

In the remainder of the paper when the symbol $A_n^{(k)}(z)$ is used k is in no way restricted to finite values. We shall understand $A_n^{(\infty)}(z) = 1/n^z$.

4.† If the Dirichlet series (3) converges when $z = z_0$, then it converges uniformly over a region defined by the inequality

$$|\text{am}(z - z_0)| \leqq \pi/2 - \epsilon_3, \quad \epsilon_3 > 0.$$

There exists a straight line $x = x_0$, where $z = x + yi$, such that when $x > x_0$ (3) converges and when $x < x_0$ it diverges. This line is called the "line of convergence."

5. If (1) converges when $z = z_0$ and $k = k_0$, then it converges uniformly in z and k over any subregion, R, of the region defined by the following inequalities

$$|z| < M, \ |\text{am}(z - z_0)| \leqq \pi/2 - \epsilon_4,$$
$$|\text{am}k| \leqq \pi - \epsilon_5, \ |k| > \epsilon_6,$$

M, ϵ_4, ϵ_5 and ϵ_6 being positive constants, which subregion is so chosen that each of the numbers $|z + k + j| > \epsilon_7 > 0$ $(j = 0, 1, 2, \ldots)$.

* See for example T. H. Gronwall, Annals of Mathematics, vol. 20 (1918–19), pp. 85, 86, 88.

† See for example G. H. Hardy and M. Riesz, *The General Theory of Dirichlet Series*, p. 3.

To prove this theorem we consider $A_n^{(k)}(z)n^z$ as given by (4). Then by the use of (5) when n is so large, say greater than n_0, that $|\operatorname{am}(nk+z)| < \pi - \epsilon_3$,

$$\frac{(nk)^z\,\Gamma(nk)}{\Gamma(nk+z)} = 1 + \frac{z-z^2}{2nk} + \frac{\psi(z,n,k)}{n^2},$$

where $|\psi| < M_2$, a constant. Moreover, again by (5), $\Gamma(z+k)/(\Gamma(k)k^z) \longrightarrow 1$ as $k \longrightarrow \infty$. And since all its finite singularities are excluded from the region in question and Γ does not vanish, it remains finite and in absolute value greater than a fixed $\epsilon_3 > 0$. Consequently for $n > n_0$,

$$A_n^{(k)}(z)n^z = f_2(z,k) + \frac{f_3(z,k)}{n} + \frac{f_4(z,k,n)}{n^2}, \tag{7}$$

where f_2, f_3 and f_4 are analytic in z. Moreover we explicitly remark that $|f_2|$, $|f_3|$, $|f_4| < M$, a constant. It can readily be shown that $1/A_n^{(k)}(z)$ is of the same form and consequently $A_n^{(k)}(z)/A_n^{(k_0)}(z)$ also, even if $k_0 \neq \infty$. For smaller values of n, if desired, a convenient definition can be given to f_2, f_3 and f_4 so that (7) will hold for all values of n.

Keep z and k restricted as above and thus assure the validity of (7). Then

$$\sum_{n=1}^{\infty} a_n A_n^{(k)}(z) = \sum_{n=1}^{\infty} a_n A_n^{(k_0)}(z)\left[f_2(z,k) + \frac{f_3(z,k)}{n} + \frac{f_4(z,k,n)}{n^2}\right]. \tag{8}$$

Consequently,* in particular, $\sum_{n=1} a_n A_n^{(\infty)}(z)$ converges at any point where (1) converges for any other value of k and conversely. Refer to Section 4, and it is immediate that the domain of z includes the half plane to the right of z_0. Moreover, again by (8) and Section 4, (1) converges uniformly in z and k.

6. A common line of convergence of $\sum_{n=1}^{\infty} a_n A_n^{(k)}(z)$ for all values of k considered follows immediately from Section 5.

7. Similarly to the results of Section 5, that is by (8), it is shown that if $\sum_{n=1}^{\infty} a_n A_n^{(k_0)}(z)$ converges absolutely then so does $\sum_{n=1}^{\infty} a_n A_n^{(k)}(z)$, z and k restricted to R.

8. The finite singularities of $F_k(z) = \sum_{n=1}^{\infty} a_n A_n^{(k)}(z)$ on the line of con-

* Well known results from the theory of infinite series.

vergence, other than a possible ones caused by a singularity of $\Gamma(z+k)$, are the same and of the same character as those of $D(z) = \sum_{n=1}^{\infty} \frac{a_n}{n^z}$.

By (4) and (5)

$$F_k(z) = \frac{k}{2+k-1} \frac{\Gamma(z+k)}{\Gamma(k)k^z} \left[D(z) + \frac{z-z^2}{k} D(z+1) + \sum_{n=1}^{\infty} \psi(z, k, n) \cdot 1/n^2 \right].$$

Here $\frac{\Gamma(z+k)}{\Gamma(k)k^z}$ and $\sum_{n=1}^{\infty} \psi(z, k, n) \cdot 1/n^2$, under the restrictions imposed on k, are analytic in z over any finite region exclusive of the singularities of $\Gamma(z+k)$, while $D(z+1)$ is analytic over a region extending one unit to the left of the line of convergence of the series defining $D(z)$. The theorem follows.

UNIVERSITY OF ALABAMA,
UNIVERSITY, ALA.

FUNCTIONS OF INFINITELY MANY VARIABLES IN HILBERT SPACE*

BY

W. L. HART

1. Introduction. A point or vector $\xi = (x_1, x_2, \ldots, x_i, \ldots)$ with the denumerably infinite number of real coördinates (x_i) is said to lie in Hilbert space if there exists for ξ a convergent modulus $M\xi$ defined by the equation

$$M\xi = \sqrt{\sum_{i=1}^{\infty} x_i^2} \quad .$$

We shall denote this space by the abbreviation "H." The present paper deals with theorems concerning real valued functions $f(\xi)$, defined for points ξ in the region R of H determined by the condition $M\xi \leqq r$, in which r is some positive number. The functions considered are in all cases assumed to be *continuous* according to a definition to be given in Section 2. In dealing with continuous functions of infinitely many variables, the range selected for ξ and the meaning assigned to the term "continuity" determine the nature of the hypotheses and proofs of the theorems which extend to infinitely many variables the various results found in the theory of functions of a real variable. In a previous paper† the author has considered functions $f(\xi)$ where ξ had as its range the region S defined by

$$|x_i - a_i| \leqq s_i \qquad (i = 1, 2, \ldots),$$

where the s_i were any positive numbers. The continuity assumption made in Paper A was so strong and the region S was of such a type that classical methods of proof with comparatively few additions sufficed for the development of a theory of real valued functions in S. In the present paper the range selected for ξ and the weak continuity assumed for the function $f(\xi)$ are of such a nature that classical methods of proof must be radically modified before they can be applied.

The point of view of the discussion below is such that, as a consequence of the Riesz-Fischer‡ Theorem concerning the Fourier constants of a summable

* Presented to the Society, September 8, 1920.

† These Transactions, vol. 18 (1917), p. 125. This article will be referred to as Paper A.

‡ Cf. Lalesco, *Équations Intégrales*, p. 95.

function, the results obtained are capable of immediate application to problems in the theory of functionals. We shall consider the analogues for functions $f(\xi)$, defined for ξ in R, of some of the most useful results in the theory of functions of a real variable. In Section 2 a mean value theorem is obtained from which there is derived a differential expression for $f(\xi)$ where the notion of a differential is defined in an appropriate manner. Hypotheses are then given under which an analogue of Taylor's theorem with an integral form for the remainder can be proved. In Section 3 the unique existence of a solution of an infinite system of ordinary differential equations is established. In Section 4 we shall consider the fundamental problem of implicit function theory for an infinite system of equations.

In Paper A results of the same character were obtained for the region S but, because of the fundamental differences mentioned above between the present viewpoint and that of Paper A, there are no essential points of contact between the theorems for the region S and those found below.

Functions in space H were first considered by Hilbert in connection with the theory of integral equations. Numerous authors, following the road opened by Hilbert, have developed the theory of linear and bilinear forms in space H together with their applications in the study of integral equations. In a recent article R. Gateaux* has considered some properties of functions $f(\xi)$ where ξ had as its range a certain type of sub-space of the region H. The present paper is not concerned with results of the type considered by Hilbert. A few of the results in Section 2 below are related to some theorems of Gateaux but no exact comparison is possible because of a fundamental difference between the region R of this paper and the range for ξ used by Gateaux. This difference will be referred to again in Section 2.

The notation used below is vectorial in character. Greek letters whenever used will denote vectors with infinitely many coördinates. Vector equations will imply the equality of the corresponding coördinates of the vectors represented by the two sides of the equality.

If $[\xi_n = (x_{1n}, x_{2n}, \ldots);\ n = 1, 2, \ldots]$ is a sequence of vectors, the limit vector, denoted by

$$\alpha = \lim_{n\to\infty} \xi_n \qquad (\alpha = a_1, a_2, \ldots),$$

is defined by the equations

$$a_i = \lim_{n\to\infty} x_{in} \qquad (i = 1, 2, \ldots)$$

provided the limits exist. A Greek letter with an added notation will represent the vector obtained by adding the given notation to each coördinate of the

*Bulletin de la Société Mathématique de France, vol. 47 (1919), p. 70.

vector denoted by the unannotated letter. For example, if $\xi = (x_1, x_2, \ldots)$, then $\xi' = (x_1', x_2', \ldots)$,

$$\xi(t) = [x_1(t), x_2(t), \ldots], \quad \frac{d\xi(t)}{dt} = \left[\frac{dx_1(t)}{dt}, \frac{dx_2(t)}{dt}, \ldots\right].$$

If ξ is in space H, this fact will be abbreviated by the notation "ξ^H." The modulus of a vector in space H will be denoted by prefixing M to the vectorial symbol. For example, if ξ and η are points in H, the modulus and the square of the modulus of $(\xi - \eta)$ are represented, respectively, by

$$M(\xi - \eta), \qquad M^2(\xi - \eta).$$

When there is no possibility of confusion arising, a sequence of quantities will be represented by the type quantity enclosed in a bracket. For example, $(\xi_n; n = 1, 2, \ldots)$ is abreviated by "(ξ_n)."

2. **Theorems on continous functions.** Let the function $f(\xi)$ be defined for a certain set R of points ξ in H. It will be convenient to use the following definition* of continuity:

DEFINITION 1. *The function $f(\xi)$ is continuous at a point α in R if, whenever*

$$\lim_{n\to\infty} M(\xi_n - \alpha) = 0, \tag{1}$$

where ξ_n is in R, it follows that

$$\lim_{n\to\infty} f(\xi_n) = f(\alpha).$$

If a sequence (ξ_n) and a point α satisfy (1) it is said that the sequence converges to α in the *strong sense*. Continuity as here defined is of a much weaker type than *complete continuity*† which is frequently postulated for functions $f(\xi)$.

Consider in the future only those points ξ^H in the region R defined by the inequality $M\xi \leqq r$. If we consider the quantity $M(\xi_n - \alpha)$ as the distance (*l'écart*) of the two points ξ_n and α, the region R is not‡ *compact*. This fact constitutes the difference already referred to between the region R and the region used by Gateaux. If a compact set of points were used instead of R a different theory would result. Some of the theorems below which would remain true in a compact region could be proved under weaker hypotheses and the proofs in certain instances would be easier. However, from the point of view of the possible applications of the present results, it seems best to choose the region R as specified above.

* Cf. Gateaux, loc. cit.

† Cf. Paper A, p. 129.

‡ Cf. Gateaux, loc. cit., p. 72.

For the region R there is the well known analogue of the Weierstrass condensation theorem which is stated in

LEMMA* 1. *If (ξ_n) is a sequence of points in R we can extract from it at least one sub-sequence (ξ_n') to which there corresponds a point α in R satisfying the equation*

$$\lim_{n\to\infty} \xi_n' = \alpha.$$

It should be noted that, as a consequence of the notational agreements of this paper, the convergence affirmed in the lemma requires merely that the coördinates (x_{in}') of ξ_n' satisfy

$$\lim_{n\to\infty} x_{in}' = a_i \qquad (\alpha = a_1, a_2, \ldots; i = 1, 2, \ldots).$$

Another well known property of the region R is considered in

LEMMA 2. *Suppose that (ξ_n) is a sequence of points in R which satisfies*

$$\lim_{n,\ m\to\infty} M(\xi_n - \xi_m) = 0. \tag{2}$$

Then, there exists a point α in R to which the sequence converges in the strong sense.

Because of its brevity the proof of this lemma will be repeated here. It is seen from (2) that, for every number $e > 0$ there exists an integer N_e such that if $n \geqq N_e$, $m \geqq N_e$, then

$$\sum_{i=1}^{k} (x_{in} - x_{im})^2 \leqq e \qquad (k = 1, 2, \ldots). \tag{3}$$

It follows from (3) and from Lemma 1 that there exists a point $\alpha = (a_1, a_2, \ldots)$ in R such that

$$\lim_{n\to\infty} \xi_n = \alpha.$$

Moreover, if m in (3) approaches infinity while n remains fixed, it is seen that we obtain the inequalities

$$\sum_{i=1}^{k} (x_{in} - a_i)^2 \leqq e \qquad (k = 1, 2, \ldots). \tag{4}$$

Hence, if $n \geqq N_e$, it follows from (4) that $M^2(\xi_n - \alpha) \leqq e$, which shows that equation (1) holds for α and the sequence (ξ_n).

It should be noted that a function $f(\xi)$, which is continuous in R, does not necessarily attain its bounds if it is bounded in R. One property of ordinary continuous functions of a real variable which does persist for $f(\xi)$ is given in the obvious

* Cf. Riesz, loc. cit., p. 57.

PROPOSITION 1. *If $f(\xi)$ is continuous at a point α in R, it follows that for every number $e > 0$ a number $d_e > 0$ can be found such that if $M(\xi - \alpha) \leqq d_e$, then*

$$| f(\xi) - f(\alpha) | \leqq e. \tag{5}$$

It will be useful later to have

PROPOSITION 2. *Suppose that $f(\xi)$ is continuous at all points in R. Then, for every ξ in R, the function of u represented by $f[\xi + u(\alpha - \xi)]$, where α is in R and $0 \leqq u \leqq 1$, is continuous with respect to α, uniformly for all values of u. That is, if*

$$\lim_{n\to\infty} M(\alpha_n - \alpha_0) = 0,$$

then

$$\lim_{n\to\infty} f[\xi + u(\alpha_n - \xi)] = f[\xi + u(\alpha_0 - \xi)], \tag{6}$$

uniformly for all values of u.

Let $\eta_n(u) = \xi + u(\alpha_n - \xi)$ and suppose that the uniformity specified in (6) does not hold. Then there must exist some number $e > 0$ such that, to every integer n there corresponds a value u_n and an integer $m' \geqq n$ such that

$$| f[\eta_{m'}(u_n)] - f[\eta_0(u_n)] | > e. \tag{7}$$

In (7) it is seen that each u_n is associated with a vector $\alpha_{m'}$. Since all values (u_n) lie on $0 \leqq u \leqq 1$, it follows that we can select a sub-sequence (u_n'), converging to a value $\bar{u}$. Let (α_n') be the sub-sequence of (α_n) which is associated with the sequence (u_n'). Then, from (7) we obtain

$$|f[\eta_n'(u_n')] - f[\eta_0(u_n')]| > e \quad (\eta_n'(u_n') = \xi + u_n'(\alpha_n' - \xi);\ n = 1, 2, \ldots). \tag{8}$$

Since $\eta_0(\bar{u}) = \xi + \bar{u}(\alpha_0 - \xi)$, it follows that

$$M[\eta_0(u_n') - \eta_0(\bar{u})] = |\bar{u} - u_n'|\, M(\xi - \alpha_0), \tag{9}$$

$$\begin{aligned} M[\eta_n'(u_n') - \eta_0(\bar{u})] &= M[(\bar{u} - u_n')(\xi - \alpha_0) + u_n'(\alpha_n' - \alpha_0)], \\ &\leqq |\bar{u} - u_n'|\, M(\xi - \alpha_0) + u_n' M(\alpha_n' - \alpha_0). \end{aligned} \tag{10}$$

The expressions on the right sides of (9) and (10) both approach zero as n approaches infinity, so that, as a result of the continuity of f, it is seen that

$$\lim_{n\to\infty} f[\eta_n'(u_n')] = \lim_{n\to\infty} f[\eta_0(u_n')] = f[\eta_0(\bar{u})],$$

which contradicts (8) if n is sufficiently large. Thus the proposition has been established.

As another obvious property of continuous functions we may state that if $f_i(\xi)$ $(i = 1, 2, \ldots)$ are continuous in R and if the series $S(\xi) = \Sigma f_i(\xi)$ converges uniformly for all ξ in R, then $S(\xi)$ is continuous in R.

For a function $f(\xi)$ defined in R we shall establish an analogue of the mean value theorem in

THEOREM I. *Let $f(\xi)$ be continuous at all points in R. Suppose that there exist partial derivatives $\partial f(\xi)/\partial x_j$ $(j = 1, 2, \ldots)$ which are continuous for ξ in R and are such that, for all ξ in R, there exists*

$$\sum_{i=1}^{\infty} \left| \frac{\partial f(\xi)}{\partial x_i} \right|^2 \leqq K, \tag{11}$$

where K is some positive number. Then, for every pair of points (ξ, ξ') in R, it follows that

$$f(\xi') = f(\xi) + \sum_{i=1}^{\infty} (x_i' - x_i) \int_0^1 \frac{\partial f[\xi + u(\xi' - \xi)]}{\partial x_i} \, du. \tag{12}$$

In order to prove the theorem let us first suppose that the two points (ξ, ξ') satisfy

$$M\xi < r, \qquad M\xi' < r.$$

We notice that, for all values of u on $0 \leqq u \leqq 1$, the point $[\xi + u(\xi' - \xi)]$ is in R because

$$M[\xi + u(\xi' - \xi)] \leqq (1 - u)M\xi + uM\xi' < r.$$

Let $\zeta_n = (x_1', \ldots, x_n', x_{n+1}, \ldots)$. Then, since ζ_n differs from ξ' only in the coördinates after the nth, it is seen that $M\zeta_n < r$ if n is sufficiently large, and that

$$\lim_{n \to \infty} M(\zeta_n - \xi') = 0.$$

Hence

$$\lim_{n \to \infty} f(\zeta_n) = f(\xi').$$

If n is so large that $M\zeta_n < r$, so that ζ_n is in the region R, it is easily established that the mean value theorem for the case of a finite number of variables can be applied to $f(\xi)$ with the result that

$$f(\zeta_n) = f(\xi) + \sum_{i=1}^{n} (x_i' - x_i) \int_0^1 \frac{\partial f[\xi + u(\zeta_n - \xi)]}{\partial x_i} \, du. \tag{13}$$

Let us consider the function $S(\alpha)$ defined, for all α in R, by the equation

$$S(\alpha) = \sum_{i=1}^{\infty} (x_i' - x_i) \int_0^1 \frac{\partial f[\xi + u(\alpha - \xi)]}{\partial x_i} \, du. \tag{14}$$

Let us first prove that $S(\alpha)$ converges uniformly for all α in R. As a consequence of the Lagrange-Cauchy* inequality for sums it is seen that the sum of the terms in (14) for $(i = n, \ldots, m)$ is, in absolute value, at most

$$\sqrt{\sum_{i=n}^{m} (x_i' - x_i)^2} \sqrt{\sum_{i=n}^{m} \left| \int_0^1 \frac{\partial f[\xi + u(\alpha - \xi)]}{\partial x_i} du \right|^2}. \tag{15}$$

As a result of the Schwarz inequality for integrals,

$$\left| \int_a^b f(x)\, g(x)\, dx \right|^2 \leqq \left| \int_a^b f^2(x)\, dx \right| \cdot \left| \int_a^b g^2(x)\, dx \right|, \tag{16}$$

it follows that (15) is at most equal to

$$\sqrt{\sum_{i=n}^{m} (x_i' - x_i)^2} \sqrt{\int_0^1 \sum_{i=n}^{m} \left| \frac{\partial f[\xi + u(\alpha - \xi)]}{\partial x_i} \right|^2 du} \leqq \sqrt{K} \sqrt{\sum_{i=n}^{m} (x_i' - x_i)^2},$$

which, as n and m approach infinity, approaches zero uniformly for all α in R. Consequently $S(\alpha)$ converges uniformly.

In order to prove that $S(\alpha)$ is a continuous function of α it is necessary for us to show that each term in the series (14) is continuous. To do this we note that, if

$$\lim_{n \to \infty} M(\alpha_n - \alpha) = 0,$$

then, as a consequence of Proposition 2,

$$\lim_{n \to \infty} \frac{\partial f[\xi + u(\alpha_n - \xi)]}{\partial x_i} = \frac{\partial f[\xi + u(\alpha - \xi)]}{\partial x_i},$$

uniformly for all values of u on (0,1). From this it follows that

$$\lim_{n \to \infty} \int_0^1 \frac{\partial f[\xi + u(\alpha_n - \xi)]}{\partial x_i} du = \int_0^1 \frac{\partial f[\xi + u(\alpha - \xi)]}{\partial x_i} du,$$

which establishes the continuity of the terms in (14).

By a slight transformation of (13) we obtain

$$f(\zeta_n) + \sum_{i=n+1}^{\infty} (x_i' - x_i) \int_0^1 \frac{\partial f[\xi + u(\zeta_n - \xi)]}{\partial x_i} du = f(\xi) + S(\zeta_n). \tag{17}$$

As a consequence of the continuity of the functions S and f it is seen that the limiting form of (17) as n approaches infinity is

$$f(\xi') = f(\xi) + S(\xi'),$$

* Riesz, *Systèmes des Équations Linéaires*, p. 36.

which is the desired equation (12). It remains to show that (12) holds when $M\xi = r$ or $M\xi' = r$. Suppose that one (or both) of these equalities has occurred, for example, $M\xi = r$. Let $\xi(v)$ represent the point whose coördinates $x_i(v)$ are defined by

$$x_i(v) = x_i \qquad (i = 1, 2, \ldots; i \neq i_0),$$
$$x_{i_0}(v) = x_{i_0} - v\frac{x_{i_0}}{x_{i_0}}\Big|,$$

where i_0 is some integer such that $x_{i_0} \neq 0$. For $v > 0$, but sufficiently small, $M\xi(v) < r$ and, for the same values of v, equation (12) is valid with ξ replaced by $\xi(v)$. The right member of (12), after this replacement, is a continuous function of v for $0 \leqq v \leqq 1$ which, as $v \to 0$, reduces to the right member of (12) as it is written in the statement of the theorem. The same procedure evidently applies if $M\xi' = r$, so that the proof of the theorem is complete.

The proof of Theorem I would have been practically unchanged if, in place of condition (11), it had been assumed that the series

$$\sum_{i=1}^{\infty} \left| y_i \frac{\partial f(\xi)}{\partial x_i} \right|$$

converged uniformly for every pair of points ξ and $\eta = (y_1, y_2, \ldots)$ in R.

Before considering the notion of the total differential of a function $f(\xi)$ let us first define a linear function $l(\xi)$ as one which satisfies the following conditions:

(*a*) $l(\xi)$ is continuous for all points ξ^H.

(*b*) For all points ξ_1^H and ξ_2^H, $l(\xi_1 + \xi_2) = l(\xi_1) + l(\xi_2)$.

In order to obtain an explicit expression for $l(\xi)$ let us consider those points in Hilbert space of the form $(x_1, \ldots, x_n, 0, 0, \ldots)$, which we may denote by $(\xi)_n$. From conditions (*a*) and (*b*), as is well known, it follows that, if δ_i^H is the point with all coördinates zero except the ith, which is 1, then

$$l[(\xi)_n] = a_1x_1 + \cdots + a_nx_n \qquad (a_i = l(\delta_i)).$$

From condition (*a*) it is seen that

$$l(\xi) = \lim_{n\to\infty} l[(\xi)_n] = \sum_{i=1}^{\infty} a_ix_i. \tag{18}$$

Since (18) converges for every ξ^H, it follows* that the point $\alpha = (a_1, a_2, \ldots)$ is itself in Hilbert space, and a function $l(\xi)$ is therefore identical with some linear form in infinitely many variables.

For a function $f(\xi)$ defined in R let us define the notion of a differential as follows:

* Cf. Riesz, loc. cit., p. 47.

DEFINITION 2. *A function $f(\xi)$ has a differential $l(\xi,\xi')$ at a point ξ of R if $l(\xi,\xi')$ is a linear function of $(\xi' - \xi)$ and if there exists a function $h(\xi,\xi')$ such that, for all ξ' of R,*

$$f(\xi') = f(\xi) + l(\xi, \xi') + h(\xi, \xi')\, M(\xi' - \xi), \tag{19}$$

where, moreover,

$$\lim_{M(\xi'-\xi) \to 0} h(\xi, \xi') = 0.$$

With the aid of Theorem I let us establish

THEOREM II. *Under the hypotheses of Theorem* I *with the condition* (11) *replaced by the assumption that the series in* (11) *converges uniformly for all ξ in R, it follows that $f(\xi)$ possesses a differential $l(\xi, \xi')$. Moreover, the $l(\xi, \xi')$ and $h(\xi, \xi')$ corresponding to f are continuous functions of (ξ, ξ') at all points in R and*

$$l(\xi, \xi') = \sum_{i=1}^{\infty} (x_i' - x_i) \frac{\partial f(\xi)}{\partial x_i}. \tag{20}$$

As a consequence of equation (12) the $l(\xi, \xi')$ defined in (20) satisfies

$$\left| f(\xi') - f(\xi) - l(\xi, \xi') \right| = \left| \sum_{i=1}^{\infty} (x'_i - x_i) \int_0^1 \left\{ \frac{\partial f[\xi + u(\xi' - \xi)]}{\partial x_i} - \frac{\partial f(\xi)}{\partial x_i} \right\} du \right|, \tag{21}$$

$$\leqq M(\xi' - \xi) \sqrt{\sum_{i=1}^{\infty} \left(\int_0^1 \left\{ \frac{\partial f[\xi + u(\xi' - \xi)]}{\partial x_i} - \frac{\partial f(\xi)}{\partial x_i} \right\} du \right)^2},$$

$$\leqq M(\xi' - \xi) \sqrt{\int_0^1 \sum_{i=1}^{\infty} \left(\frac{\partial f[\xi + u(\xi' - \xi)]}{\partial x_i} - \frac{\partial f(\xi)}{\partial x_i} \right)^2 du}, \tag{22}$$

where the second inequality is a consequence of (16). In view of the type of convergence assumed for series (11), it is seen that the expression under the radical in (22) converges uniformly for all values of (u, ξ, ξ'). A course of reasoning similar to that used in regard to $S(\alpha)$ in (14) shows that the expression under the radical is a continuous function of ξ' and hence approaches zero if ξ' varies in such a way that $M(\xi' - \xi) \to 0$.

Let us now define $h(\xi, \xi')$ by the equations

$$\begin{aligned} &h(\xi, \xi) = 0, \\ &h(\xi, \xi')M(\xi' - \xi) = f(\xi') - f(\xi) - l(\xi, \xi') \quad (\xi \neq \xi'). \end{aligned} \tag{23}$$

It is easily verified that the equation (19) is satisfied by l and h as defined in (20) and (23). Moreover, since $l(\xi, \xi')$ of equation (20) is continuous in (ξ, ξ') for every pair of points in R, it follows from equation (19) that $h(\xi, \xi')$ possesses the same property. This completes the proof of Theorem II.

By precisely the same method as was used in the proof of Theorem I we obtain in Theorem III below an analogue of Taylor's theorem for $f(\xi)$. The proof will not be given since, except for the greater complication of the formulas, the steps would be almost identical with those of the proof of Theorem I.

THEOREM III. *Let $f(\xi)$ satisfy the hypotheses of Theorem I. Suppose that all partial derivatives used in the formulas below exist and are continuous for all points ξ in R. Assume that the series*

$$\sum_{i_1,\ldots,i_m=1}^{\infty} \left| y_{i_1} \ldots y_{i_m} \frac{\partial^{(m)} f(\xi)}{\partial x_{i_1} \ldots \partial x_{i_m}} \right| \qquad (m = 1, 2, \ldots, n,)$$

converge uniformly for all points $\eta = (y_1, y_2, \ldots)$ and ξ in R. Then, for every pair of points (ξ, ξ') in R, it follows that

$$f(\xi') - f(\xi) = \sum_{i=1}^{\infty} (x_i' - x_i) \frac{\partial f(\xi)}{\partial x_i} + \sum_{i_1, i_2 = 1}^{\infty} \frac{(x_{i_1}' - x_{i_1})(x_{i_2}' - x_{i_2})}{2!} \frac{\partial^2 f(\xi)}{\partial x_{i_1} \partial x_{i_2}}$$

$$+ \cdots + \sum_{i_1, \ldots, i_{n-1} = 1}^{\infty} \frac{(x_{i_1}' - x_{i_1}) \ldots (x_{i_{n-1}}' - x_{i_{n-1}})}{(n-1)} \frac{\partial^{(n-1)} f(\xi)}{\partial x_{i_1} \ldots \partial x_{i_{n-1}}}$$

$$+ \sum_{i_1, \ldots, i_n = 1}^{\infty} \frac{(x_{i_1} - x_{i_1}) \ldots (x_{i_n}' - x_{i_n})}{(n-1)!} \int_0^1 \frac{\partial^{(n)} f[\xi + u(\xi' - \xi)]}{\partial x_{i_1} \ldots \partial x_{i_n}} (1-u)^{n-1} du.$$

3. Solution of an infinite System of differential equations. Consider the infinite system of ordinary differential equations

$$\frac{dx_i}{dt} = f_i(t, \xi) \qquad (\xi = |x_1, x_2, \ldots|; i = 1, 2, \ldots), \tag{24}$$

which in vector notation becomes

$$\frac{d\xi}{dt} = \phi(t, \xi) \qquad (\phi = f_1, f_2, \ldots).$$

Systems of the form (24) in which the f_i were linear functions of ξ entered as special cases in certain theories of E. H. Moore* and of T. H. Hildebrandt† in the domain of Moore's *General Analysis*. The author, also, has considered‡ a theory of linear systems in which the right members of the equations were analytic functions of t. Certain of the linear systems included under the general theories of Moore and of Hildebrandt, as well as the linear type consid-

* *Atti del IV Congresso Internazionale dei Matematici* (Roma, 1908), vol. 2, p. 98.

† These Transactions, vol. 18 (1917), p. 73.

‡ American Journal of Mathematics, vol. 39 (1917), p. 407.

ered previously by the author, when restricted to real values of the quantities entering, would satisfy the hypotheses imposed below on the system (24). In a course of lectures, which the author attended at the University of Chicago in 1915, E. H. Moore* established the existence of the solution of a differential equation of a general form in the field of General Analysis. The postulates of Moore, when interpreted for (24), differ in certain essential features from the hypotheses under which the system is considered in this paper.

In the consideration of (24) it will be convenient to denote by W the class of all functions $\xi(t)$ with the following properties:

(*a*) For every value of t on some interval $|t| \leqq T_1$ the function $\xi(t)$ is in R.

(*b*) If $|t'| \leqq T_1$ and $|t''| \leqq T_1$, then

$$\lim_{|t'-t''|\to 0} M[\xi(t') - \xi(t'')] = 0.$$

The existence of a solution of (24) will be established in

THEOREM IV. *Let the functions f_i in (24) be defined and continuous simultaneously in their arguments for $|t| \leqq T$, and ξ in R. Assume that for every (t, ξ) the modulus $M\,\phi(t, \xi)$ exists and that there is a number $A > 0$ such that, for all $|t| \leqq T$,*

$$M[\phi(t, \xi_1) - \phi(t, \xi_2)] \leqq AM(\xi_1 - \xi_2), \tag{25}$$

for every pair of points (ξ_1, ξ_2) in R. Suppose, furthermore, that $M\phi(t, 0)$ has an upper bound $b > 0$ for all $|t| \leqq T$. Then, if $|t|$ is sufficiently small there exists, among functions of the class W, one and only one function $\xi(t)$, satisfying (24) and the initial condition $\xi(0) = 0$.

Let us first note that there exists a number $B > 0$ such that, for all (t, ξ),

$$M\phi(t, \xi) \leqq B. \tag{26}$$

This fact is a consequence of the inequality

$$\begin{aligned} M\phi(t, \xi) &\leqq M[\phi(t, \xi) - \phi(t, 0)] + M\phi(t, 0) \\ &\leqq AM\xi + M\phi(t, 0) \leqq Ar + b, \end{aligned}$$

from which it follows that (26) is true with $B = Ar + b$.

We also note that, if $\xi(t)$ is in the class W for $|t| \leqq T_1 \leqq T$, the expressions $f_i[t, \xi(t)]$ are continuous in t because, if $\lim_{n\to\infty} t_n = \bar{t}$, it follows that

$$\lim_{n\to\infty} M[\xi(t_n) - \xi(\bar{t})] = 0,$$

so that

$$\lim_{n\to\infty} f_i[t_n, \xi(t_n)] = f_i[\bar{t}, \xi(\bar{t})].$$

* These lectures are unpublished.

We shall show the existence of the solution $\xi(t)$ by proving that such a function is approached in the strong sense by the sequence $[\xi_k^{(t)}]$ defined formally by the equations

$$\xi_0 = 0$$
$$\xi_k(t) = \int_0^t \phi[t, \xi_{k-1}(t)]dt \qquad (k = 1, 2, \ldots). \tag{27}$$

For the sake of simplicity in certain steps of the proof suppose that $t \geqq 0$; obvious modifications would suffice for the case $t < 0$. In all integrals occurring below, the integration is with respect to t and, for convenience, the symbol dt will be omitted in the integrands. The convergence of the sequence $[\xi_k(t)]$ will be established in two steps. It will first be shown that a number $d > 0$ exists so that for $|t| \leqq d$, the functions of (27) are defined and belong to the class W. It will then be shown that, for $|t| \leqq d$, the sequence (27) converges in the strong sense to a solution $\xi(t)$.

To accomplish the first step mentioned we shall show that there exists a number $d > 0$, $d < T$, such that, for every number $e > 0$ there can be found a number $g_e > 0$ such that, for all values of k and for all values of (t, t') satisfying

$$|t - t'| \leqq g_e, \quad |t| \leqq d, \quad |t'| \leqq d,$$

we have the inequalities.

$$M[\xi_k(t) - \xi_k(t')] \leqq e, \quad M\xi_k(t) \leqq r. \tag{28}$$

Consider (28) for $k = 1$. With the aid of (16) it is easily verified that

$$\sum_{i=1}^{n}\left[\int_t^{t'} f_i(t,0)\right]^2 \leqq |t-t'| \sum_{i=1}^{n}\left|\int_t^{t'} f_i^2(t,0)\right| = |t-t'| \left|\int_t^{t'} \sum_{i=1}^{n} f_i^2(t,0)\right|, \tag{29}$$
$$\leqq |t-t'| \left|\int_t^{t'} M^2\phi(t,0)\right| \leqq B^2|t-t'|^2.$$

Since (29) is true for all values of n it is seen that

$$M^2\left[\xi_1(t') - \xi_1(t)\right] = \sum_{i=1}^{\infty}\left[\int_t^{t'} f_i(t,0)\right]^2 \leqq B^2|t-t'|^2. \tag{30}$$

Let d be the smaller of T and r/B. Then it follows from (30) that (28) is true for $k = 1$, so that $\xi_1(t)$ is a function of the class W. Consequently, the expressions $f_i[t, \xi_1(t)]$ are continuous in t and therefore $\xi_2(t)$, as defined in (27), exists. The same procedure as in (29) and (30) now suffices to establish the truth of (28) for all values of k. The number g_e noted above in connection with (28) can be selected independent of k because the right member of the inequality corresponding to (30) would be the same for the function $\xi_k(t)$, $k > 1$, as it is for $\xi_1(t)$.

To perform the second step of the proof let us first show by induction that

$$M\left[\xi_k(t)-\xi_{k-1}(t)\right] \leqq \frac{A^{k-1}B^k t^k}{\sqrt{k!}} \qquad (k = 1, 2, \ldots). \tag{31}$$

From the defining equations (27) and from (30) with $t' = 0$, we obtain the equations

$$\begin{aligned} M[\xi_1(t)-\xi_0] &\leqq Bt, \\ M^2[\xi_2(t)-\xi_1(t)] &= M^2\left[\int_0^t \phi[t, \xi_1(t)]-\phi(t,0)\right] \\ &= \sum_{i=1}^{\infty}\left[\int_0^t f_i[t, \xi_1(t)]-f_i(t,0)\right]^2. \end{aligned} \tag{32}$$

As a consequence of (25) it is seen that

$$\begin{aligned} M^2\left[\phi[t, \xi_1(t)]-\phi(t,0)\right] &= \sum_{i=1}^{\infty}\left[f_i[t, \xi_1(t)]-f_i(t,0)\right]^2 \leqq A^2M^2(\xi_1(t)-\xi_0) \\ &\leqq A^2B^2t^2. \end{aligned} \tag{33}$$

Since, as a consequence of (16), the sum in (32) for $i = 1, \ldots, n$ is at most

$$\begin{aligned} t\left(\sum_{i=1}^{n}\int_0^t\left\{f_i[t, \xi(t)]-f_i(t,0)\right\}^2\right) &= t\int_0^t\sum_{i=1}^{n}\left(f_i[t, \xi_1(t)]-f_i(t,0)\right)^2 \\ &\leqq t\,\frac{(A^2B^2t^3)}{3}, \end{aligned} \tag{34}$$

it follows that

$$M^2\left[\xi_2(t)-\xi_1(t)\right] \leqq \frac{t(A^2B^2t^3)}{3} \leqq \frac{A^2B^2t^4}{2!}$$

which establishes (31) for the case $k = 2$. A simple induction proof, in which the steps would be the same as those taken in (32) to (34), would now suffice to establish (31) for all values of k.

As a consequence of (31) for $k = m + 1, \ldots, n$, it follows that

$$M\left[\xi_n(t)-\xi_m(t)\right] \leqq \sum_{k=m+1}^{n}\frac{A^{k-1}B^k t^k}{\sqrt{k!}} \qquad (n > m). \tag{35}$$

Since the series, whose general term is the expression under the summation sign in (35), converges uniformly for $|t| \leqq d$, it is seen that the left member of (35) approaches zero as n and m approach infinity, uniformly for $|t| \leqq d$. As a consequence of this uniformity it is readily established, by a method similar to that used in Lemma 2, §2, that there exists a function $\xi(t)$ such that

$$\lim_{n\to\infty} M[\xi_n(t)-\xi(t)] = 0,$$

uniformly for $|t| \leqq d$. Moreover, since the g_ϵ related to (28) was independent of k, it follows that

$$\lim_{|t-t'|\to 0} M[\xi(t')-\xi(t)]=0, \tag{36}$$

uniformly for all $|t| \leqq d$, $|t'| \leqq d$. Thus $\xi(t)$ is a function of the class W.

To show that $\xi(t)$ is a solution of (24) it is first necessary to prove that, for every value of i,

$$\lim_{n\to\infty} f_i[t, \xi_n(t)] = f_i[t, \xi(t)], \tag{37}$$

uniformly for $|t| \leqq d$. Let us establish this by an indirect argument. Suppose, therefore, that the uniformity does not hold. It follows, by reasoning such as was used in Proposition 2, that a sequence of points (t_n) can be found on the interval $|t| \leqq d$ with $\lim_{n\to\infty} t_n = \bar{t}$ to which there corresponds a sub-sequence $\{\xi'_n(t)\}$ of the functions of (27), satisfying the equations

$$|f_i[t_n, \xi'_n(t_n)]-f_i[t_n, \xi(t_n)]| > e \qquad (n = 1, 2, \ldots). \tag{38}$$

In view of the fact that

$$M[\xi'_n(t_n)-\xi(t_n)] \leqq M[\xi'_n(t_n)-\xi'_n(\bar{t})] + M[\xi'_n(\bar{t})-\xi(\bar{t})] + M[\xi(\bar{t})-\xi(t_n)],$$

which approaches zero as n approaches infinity, it follows from the continuity of $f_i(t, \xi)$ that the left side of (38) approaches zero as n becomes infinite. Hence (38) has been contradicted and uniformity must exist in (37). Because of this uniformity it is seen that the limiting form of (27) as k approaches infinity is

$$\xi(t) = \int_0^t \phi[t, \xi(t)],$$

from which (24) is obtained by differentiation.

It remains to prove that $\xi(t)$ is the unique solution of (24). Suppose there were a second solution $\eta(t)$ belonging to the class W for $|t| \leqq d_1 \leqq d$. Then $M[\xi(t) - \eta(t)] \leqq 2r$. Moreover, we verify that

$$\begin{aligned} M^2[\xi(t)-\eta(t)] &= \sum_{i=1}^{\infty} \left\{ \int_0^t f_i[t, \xi(t)]-f_i[t, \eta(t)] \right\}^2 \\ &\leqq |t| \sum_{i=1}^{\infty} \left| \int_0^t \left\{ f_i[t, \xi(t)]-f_i[t, \eta(t)] \right\}^2 \right| \\ &\leqq |t|^2(4r^2A^2), \end{aligned}$$

where the last inequality follows from (25) by the procedure illustrated in (34). In like manner we obtain in succession

$$M^2[\xi(t)-\eta(t)] \leqq \frac{4r^2A^{2k}|t|^{2k}}{k!} \qquad (k = 1, 2, \ldots),$$

from which it is seen that $\eta(t)=\xi(t)$ for $|t| \leqq d_1$. This completes the proof of Theorem IV.

With the assistance of Theorem I it is easy to establish

COROLLARY 1. *In place of condition* (25), *assume that there exist partial derivatives* $\partial f_i/\partial x_j$ *which are continuous in* (t, ξ) *and are such that, for all* (t,ξ) *satisfying* $M\xi \leqq r$, $|t| \leqq T$, *we have*

$$\left|\frac{\partial f_i(t, \xi)}{\partial x_j}\right| \leqq b_{ij},$$

where $\Sigma_{ij=1}^{\infty}\, b_{ij}^2$ *converges. Then, there exists a constant* A *satisfying* (25).

Because of Theorem I we may state that

$$\left|f_i(t, \xi_1)-f_i(t, \xi_2)\right|^2 = \left|\sum_{j=1}^{\infty}(x_{j_1}-x_{j_1}) \int_0^1 \frac{\partial f_i[t, \xi_1 + u(\xi_2-\xi_1)]}{\partial x_j}\, du\right|^2$$

$$\leqq M^2(\xi_2-\xi_1)\sum_{ij=1}^{\infty} b_{ij}^2 .$$

Hence (25) is satisfied with $A = \sqrt{\sum_{ij=1}^{\infty} b_{ij}^2}$.

In the next corollary there is established a characteristic of the convergence of the difference quotients $\Delta x_i(t)/\Delta t$ to the derivatives $dx_i(t)/dt$ which is useful in an application of Theorem IV to be treated in a later paper.

COROLLARY 2. *In addition to the hypotheses of Theorem* IV *assume that there exists a number* $C > 0$ *such that, for* ξ *in* R *and* $|t_1| \leqq T$, $|t_2| \leqq T$,

$$M[\phi(t_1, \xi)-\phi(t_2, \xi)] \leqq C|t_1-t_2|. \tag{39}$$

Then, the function $\xi(t)$ *satisfies the equation*

$$\lim_{\Delta t \to 0} M\left[\frac{\xi(t+\Delta t)-\xi(t)}{\Delta t}-\phi[t, \xi(t)]\right] = 0, \tag{40}$$

uniformly for all values $|t| \leqq d$.

From equation (24) it is seen that, if $|t_0| \leqq d$,

$$\frac{\xi(t_0+\Delta t)-\xi(t_0)}{\Delta t}-\phi[t_0, \xi(t_0)] = \int_{t_0}^{t_0+\Delta t}\left\{\frac{\phi[t, \xi(t)]-\phi[t_0, \xi(t_0)]}{\Delta t}\right\}.$$

Call the quantity on the left $\eta(t_0, \Delta t)$. It is easily verified that

$$(\Delta t)^2 M^2 \eta(t_0, \Delta t) = \sum_{1=i}^{\infty}\left[\int_{t_0}^{t_0+\Delta t}\left\{f_i[t, \xi(t)]-f_i[t_0, \xi(t_0)]\right\}\right]^2. \tag{41}$$

The sum of the terms on the right in (41) for $i = 1, \ldots, n$ is at most

$$(42)\quad 2\,|\,\Delta t\,|\left|\int_{t_0}^{t_0+\Delta t}\sum_{i=1}^{n}\left\{f_i[t,\xi(t)]-f_i[t_0,\xi(t)]\right\}^2+\left\{f_i[t_0,\xi(t)]-f_i[t_0,\xi(t_0)]\right\}^2\right|.$$

If Δt is so small that, as a consequence of (36), $M^2[\xi(t)-\xi(t_0)] \leqq e$ for $t_0 \leqq t \leqq t_0+\Delta t$, it follows from (25), (39) and (42) that

$$(43)\quad \begin{aligned}\Delta t^2 M^2\eta(t_0,\Delta t) &\leqq 2\,|\,\Delta t\,|\left|\int_{t_0}^{t_0+\Delta t}\left[C^2(\Delta t)^2+A^2e\right]\right| \\ &\leqq 2(\Delta t)^2\left[C^2(\Delta t)^2+A^2e\right].\end{aligned}$$

Since e in (43) can be selected arbitrarily small it follows that (40) holds for $t = t_0$. The uniformity affirmed in the corollary is a consequence of the fact that the steps in the reasoning above can be accomplished uniformly for all $|\,t_0\,| \leqq d$.

4. Implicit functions. Let us consider the solution of the infinite system of equations

$$f_i(t,\xi)=0 \qquad (\xi=x_1,x_2,\ldots;i=1,2,\ldots),$$

which, in the vector notation, become

$$(44)\qquad \phi(t,\xi)=0 \qquad (\phi=f_1,f_2,\ldots),$$

where (t,ξ) are in the region P defined by

$$|\,t\,| \leqq T, \qquad M\xi \leqq r.$$

Suppose that $(t = 0, \xi = 0)$ satisfies (44). We shall establish the existence of a solution $\xi(t)$ for which $\xi(0)=0$, in

THEOREM V. *In system* (44) *assume that the* $f_i(t,\xi)$ *are continuous in their arguments if* (t,ξ) *are in* P. *Suppose that for all points in* P *the series* $\Sigma_{i=1}^{\infty} f_i^2(t,\xi)$ *converges uniformly. Assume, moreover, that all derivatives* $\partial f_i(t,\xi)/\partial x_j$ $(i, j = 1, 2, \ldots)$ *exist and are continuous in* P *and that the two series*

$$(45)\qquad \sum_{i=1}^{\infty}\left|1-\frac{\partial f_i(t,\xi)}{\partial x_i}\right|, \qquad \sum_{(i,j=1,\ i\neq j)}^{\infty}\left|\frac{\partial f_i(t,\xi)}{\partial x_j}\right|^2,$$

converge uniformly for (t,ξ) *in* P. *Then, if the convergent infinite determinant*

$$\Delta=\left|A_{ij}\right|_{i,j=1,2,\ldots}\neq 0 \qquad \left[A_{ij}=\frac{\partial f_i(0,0)}{\partial x_j}\right],$$

it follows that there exist positive numbers $c \leqq 1$, $d \leqq 1$ *such that, for all points* t *on the interval* $|\,t\,| \leqq cT$, *there corresponds one and only one point* $\xi(t)$ *in the region*

$M\xi \leqq dr$ such that $[t, \xi(t)]$ satisfies (44). *All coördinates of the function $\xi(t)$ so determined are continuous for $|t| \leqq cT$.*

The hypotheses of Theorem V as stated are stronger than is necessary and have been made so for convenience in the proof. At the end of the present section weaker hypotheses leading to the same conclusions will be mentioned. The theorem will be established by first obtaining a system

$$\xi = \psi(t, \xi) \qquad (\psi = h_1, h_2, \ldots), \tag{46}$$

equivalent to (44) and by then solving (46) by the method of successive substitutions.

Define $\psi(t, \xi)$ as the function whose coördinates $h_i(t, \xi)$ are given by the solution of the infinite system of linear equations.

$$f_i(t, \xi) + \sum_{j=1}^{\infty} A_{ij}(h_j - x_j) = 0 \qquad (i = 1, 2, \ldots). \tag{47}$$

Let $A_{ij} = d_{ij} + a_{ij}$ where $d_{ii} = 1$, $d_{ij} = 0$ $(i \neq j)$. The infinite system (47) is of a type* for which the determinant Δ and all of its minors converge absolutely. Since we have assumed $\Delta \neq 0$, it follows† that there exists an infinite matrix

$$B = (b_{ij})_{i,j=1,2,\ldots}$$

which is the reciprocal‡ of the matrix $(-A_{i,j})_{ij=1,2,\ldots}$ and by means of which the solution of (47) is given in the form

$$h_j(t, \xi) - x_j = \sum_{k=1}^{\infty} b_{jk} f_k(t, \xi) \qquad (j = 1, 2, \ldots) \tag{48}$$

Moreover, since $\Delta \neq 0$, the system (44) is equivalent§ to the system (46) with $\psi(t, \xi)$ defined by (47). After certain preliminary considerations we shall solve (46) by the method of successive substitutions.

Let us note some properties of the b_{ij} which result from the special character of (47) and are consequences of the theorems proved in the sections of *Equations Linéaires* referred to in the preceding paragraph. Since, for every point $\alpha = (a_1, a_2, \ldots)$ in space H, the system

$$\sum_{j=1}^{\infty} A_{ij} z_j = a_i \qquad (i = 1, 2, \ldots) \tag{49}$$

---

* Riesz, loc. cit., p. 35.

† Riesz, loc. cit., §§ 30, 68, 69.

‡ Hellinger-Toeplitz, Mathematische Annalen, vol. 69 (1910), p. 311. In the future, references to this article will be embodied in the text by the notation "H. T. p. (—)."

§ Cf. Riesz, loc. cit., p. 29.

has a unique solution $\zeta = (z_1, z_2, \ldots)$ in H, where

$$z_j = -\sum_{k=1}^{\infty} b_{jk} a_k \qquad (j = 1, 2, \ldots), \tag{50}$$

it follows (H.T. p. 321) that the matrix B is *limited* in the Hilbert sense (H.T. p. 296). That is, there exists a bound $b > 0$ such that, for all points ξ^n and $\eta^n = (y_1, y_2, \ldots)$,

$$\left|\sum_{i,j=1}^{n} x_i b_{ij} y_j\right| \leqq bM\xi M\eta \qquad (n = 1, 2, \ldots). \tag{51}$$

Since the matrix B is *limited* with the bound b it follows (H.T. p. 297) that each of the series

$$\left(\sum_{k=1}^{\infty} b_{jk}^2;\ j = 1,\ 2,\ \ldots\ \right),\quad \left(\sum_{j=1}^{\infty} b_{jk}^2;\ k = 1,\ 2,\ \ldots\ \right)$$

converges and is at most b^2. Moreover, for every point $\alpha = (a_1, a_2, \ldots)$ in H, the point ζ whose coördinates z_j are given by (50) satisfies (H.T. p. 301) the equation

$$M\zeta \leqq bM\alpha. \tag{52}$$

On using (52) for the case of (47), we obtain

$$M[\psi(t, \xi) - \xi] \leqq bM\phi(t, \xi). \tag{53}$$

It should be noted that, as a consequence of the present hypotheses, $M\phi(t, \xi)$ is a continuous function of (t, ξ), and, moreover, $M\phi(0, 0) = 0$.

It will be necessary later to use the derivatives $\partial h_i/\partial x_j$ which, provided that the series used below converge uniformly, may be written

$$\frac{\partial h_i}{\partial x_j} = d_{ij} + \sum_{k=1}^{\infty} b_{ik} \frac{\partial f_k(t, \xi)}{\partial x_j} \qquad (i, j = 1, 2, \ldots),$$

$$\frac{\partial h_i}{\partial x_j} = \sum_{k=1}^{\infty} b_{ik}\left(\frac{\partial f_k(t, \xi)}{\partial x_j} - A_{ki}\right). \tag{54}$$

The last reduction in (54) was a consequence of the equations

$$d_{ij} = \sum_{k=1}^{\infty} -b_{ik} A_{kj} \qquad (i, j = 1, 2, \ldots)$$

which hold because $-B$ is the reciprocal of the matrix (A_{ij}). The uniform convergence needed in writing (54) is established in

LEMMA 3. *The infinite series*

$$S(t, \xi) = \sum_{i,k=1}^{\infty} \left|\frac{\partial f_k(t, \xi)}{\partial x_i} - A_{ki}\right|^2 \tag{55}$$

converges uniformly for all (t, ξ) *in* P *and, as a consequence, the same fact can be stated concerning each of the series* (54). *Moreover,* $S(t, \xi)$ *is continuous if* (t, ξ) *are in* P.

By an obvious transformation of (55) we obtain the inequality

$$S(t,\xi) \leqq 2\sum_{i,k=1}^{\infty} \left|\frac{\partial f_k(t,\xi)}{\partial x_i} - d_{ki}\right|^2 + 2\sum_{i,k=1}^{\infty} \left|d_{ki} - A_{ki}\right|^2.$$

The uniform convergence of the series on the right, and hence the uniform convergence of $S(t, \xi)$, is implied by the type of convergence assumed for the series (45). The uniform convergence of (54) can be inferred from the inequality

$$\left|\sum_{k=n}^{m} b_{ik}\left(\frac{\partial f_k(t,\xi)}{\partial x_j} - A_{kj}\right)\right|^2 \leqq b^2 \sum_{k=n}^{m}\left(\frac{\partial f_k(t,\xi)}{\partial x_j} - A_{kj}\right)^2, \tag{56}$$

which is obtained by use of the Lagrange-Cauchy inequality. When n and m approach infinity the right member of (56) approaches zero uniformly for (t, ξ) in P because of the uniformity in (55). The continuity of $S(t, \xi)$ is evident because each term in the series (55) is continuous.

It is seen that $S(0, 0) = 0$. Hence, *if a positive number* $w < 1$ *is assigned, a positive number* $d < 1$ *can be found so that* $S(t, \xi) \leqq w^2/b^2$ *if* (t, ξ) *satisfy* $M\xi \leqq dr$, $|t| \leqq dT$.

With the aid of the properties possessed by $S(t, \xi)$, let us show that a positive number $c \leqq d$ can be selected so small that c and d satisfy the statement of Theorem V. In order to do this let us consider the sequence $[\xi_k(t)]$ defined by the equations

$$\begin{aligned} \xi_0 &= 0 \\ \xi_k(t) &= \psi[t, \xi_{k-1}(t)] \end{aligned} \qquad (k = 1, 2, \ldots). \tag{57}$$

We shall first prove that a number $c \leqq d$ can be found so that for all values $|t| \leqq cT$ the sequence (57) is defined and satisfies $M\xi_k(t) \leqq dr$. As a consequence of (53) we obtain the inequality

$$M\xi_1(t) \leqq bM\phi(t, 0).$$

Since $M\phi(0, 0) = 0$, we can select $c \leqq d$ so that, if $|t| \leqq cT$,

$$M\xi_1(t) \leqq dr(1-w) < dr. \tag{58}$$

By use of Theorem I we derive the equation

$$\begin{aligned} \xi_2(t) - \xi_1(t) &= \psi[t, \xi_1(t)] - \psi(t, 0) \\ &= \sum_{j=1}^{\infty} x_{j1}(t) \int_0^1 \frac{\partial \psi[t, \xi_1(t) + u(\xi_1(t) - \xi_0)]}{\partial x_j}\, du. \end{aligned} \tag{59}$$

The coordinates of the derivatives $\partial\psi/\partial x_j$ are given by (54) and it is seen that the condition (11) of Theorem I is satisfied in the present case because, if $|t| \leqq cT$ and $M\xi \leqq dr$,

$$\sum_{j=1}^{\infty}\left(\frac{\partial h_i(t,\xi)}{\partial x_j}\right)^2 \leqq b^2 S(t,\xi) \leqq \frac{b^2w^2}{b^2} = w^2 \qquad (i=1,2,\ldots).$$

LEMMA 4. *If* $|t| \leqq cT$, $M\xi' \leqq dr$, $M\xi'' \leqq dr$, *the infinite matrix* $K = (k_{ij})$, *whose elements are given by*

$$k_{ij} = \int_0^1 \frac{\partial h_i[t, \xi' + u(\xi''-\xi')]}{\partial x_j}\,du \qquad (i,j=1,2,\ldots),$$

is limited and its bound may be taken as w.

First consider the matrix

$$F = \left(\frac{\partial h_i(t,\xi)}{\partial x_j}\right)_{i,j=1,2,\ldots}.$$

From equation (54) it is seen that F is the product of the matrices B and G where

$$G = \left(\frac{\partial f_k(t,\xi)}{\partial x_j} - A_{kj}\right)_{k,j=1,2,\ldots}.$$

If $|t| \leqq cT$ and $M\xi \leqq dr$, it follows (H. T. p. 307) that G is limited and that its bound may be taken as $\sqrt{S(t,\xi)} \leqq w/b$. Hence the matrix F is limited and has the bound $b(w/b) = w$ (H. T. p. 300). To show that K has the bound w it is merely necessary to note that, if $\alpha = (a_1, a_2, \ldots)$ and $\beta = (b_1, b_2, \ldots)$ are in Hilbert space then, on account of the limited character of F,

$$(60) \qquad \left|\sum_{ij=1}^{n} a_i k_{ij} b_j\right| = \left|\int_0^1 \sum_{ij=1}^{n} a_i b_j \frac{\partial h_i[t, \xi' + u(\xi''-\xi')]}{\partial x_j}\,du\right| \leqq wM\alpha M\beta.$$

From the definition of a limited matrix in (51) it is seen that (60) completes the proof of the lemma.

In order to establish Theorem V, return to equation (59). From Lemma 4 and from equation (52), interpreted for the matrix K instead of B, it follows that

$$M[\xi_2(t) - \xi_1(t)] \leqq wM\xi_1(t) \leqq wrd(1-w) < dr.$$

Hence it is seen that

$$M\xi_2(t) \leqq M[\xi_2(t)-\xi_1(t)] + M\xi_1(t) \leqq dr(1-w)(1+w) = dr(1-w^2).$$

In a similar manner it can be proved by induction that

$$(61) \qquad M[\xi_k(t) - \xi_{k-1}(t)] \leqq wM[\xi_{k-1}(t) - \xi_{k-2}(t)] \leqq w^{k-1}dr(1-w).$$

From equation (61) it is easily established by the method used in connection with (35) in Theorem IV, that there exists a function $\xi(t)$ such that

$$\lim_{k\to\infty} M[\xi_k(t)-\xi(t)]=0,$$

uniformly for $|t| \leqq cT$. On taking the limiting form of (57) as k approaches infinity, we obtain

$$\xi(t)=\psi[t, \xi(t)],$$

and hence $\xi(t)$ is also a solution of (44).

The solution obtained is the only one existing for $|t| \leqq cT$ because, if for some value $|t_1| \leqq cT$ there was a second solution $\eta^H=(y_1, y_2, \ldots) \neq \xi(t_1)$ and satisfying $M\eta \leqq dr$, then

$$\begin{aligned}\xi(t_1)-\eta&=\psi[t_1, \xi(t_1)]-\psi(t_1, \eta)\\&=\sum_{j=1}^{\infty}(x_j(t_1)-y_j)\int_0^1 \frac{\partial\psi[t_1, \eta+u(\xi(t_1)-\eta)]}{\partial x_j}\,du.\end{aligned}$$

From this it follows from Lemma 4 that

$$M[\xi(t_1)-\eta] \leqq wM[\xi(t_1)-\eta],$$

which is a contradictory statement unless $\xi(t_1)=\eta$. This completes the proof of Theorem V.

Let us consider how the hypotheses of Theorem V could be weakened. Instead of postulating the uniform convergence of $\sum_{i=1}^{\infty} f_i^2(t, \xi)$ it could be assumed that each function $f_i(t, \xi)$ has an upper bound $B_i > 0$. We could then obtain a system, equivalent to (44) and satisfying the conditions of Theorem V, by dividing each function $f_i(t, \xi)$ by iB_i.

In the second place, by use of certain results from the theory of infinite determinants,* Theorem V could be proved with the first condition of (45) replaced by the assumption that

$$\sum_{i=1}^{\infty}\left(1-\frac{\partial f_i(t, \xi)}{\partial x_i}\right)^2$$

converges uniformly.

It is not intended in the present paper to consider applications of the results that have been obtained. It should be noted, however, that the solution of the system of differential equations (24) suggests a means for the solution of a related type of differential-functional equations. The solution of the system (44) suggests a method for solving a related type of functional equations. The author will consider these applications in detail in a later paper.

* Cf. Riesz, loc. cit., p. 39.

UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.

PRIME AND COMPOSITE POLYNOMIALS*

BY

J. F. RITT

I. INTRODUCTION

Let $F(z)$ represent any polynomial in z of degree greater than unity.† If there exist two polynomials, $\varphi_1(z)$ and $\varphi_2(z)$, each of degree greater than unity, such that

$$F(z) = \varphi_1[\varphi_2(z)], \tag{1}$$

we shall say that $F(z)$ is *composite*. If no such pair of polynomials exists we shall say that $F(z)$ is *prime*.‡

It will be convenient to omit the variable z and all symbols of aggregation from our notations, writing (1), for instance, $F = \varphi_1\varphi_2$.

If we have

$$F = \varphi_1 \varphi_2 \dots \varphi_r \tag{2}$$

where each $\varphi_i(z)$ is a polynomial of degree greater than unity, prime or composite, we shall say that (2) is a *decomposition* of $F(z)$.

The first result of the present paper is that *any two decompositions of a given polynomial into prime polynomials contain the same number of polynomials; the degrees of the polynomials in one decomposition are the same as those in the other, except, perhaps, for the order in which they occur.*

Two decompositions of $F(z)$ into the same number of polynomials,

$$F = \varphi_1 \varphi_2 \dots \varphi_r, \qquad F = \psi_1 \psi_2 \dots \psi_r,$$

will be said to be *equivalent* if there exist $r-1$ polynomials of the first degree

$$\lambda_1(z), \lambda_2(z), \dots, \lambda_{r-1}(z)$$

such that

$$\psi_1 = \varphi_1\lambda_1, \quad \psi_2 = \lambda_1^{-1}\varphi_2\lambda_2, \quad \dots, \quad \psi_r = \lambda_{r-1}^{-1}\varphi_r.$$

* Presented to the Society, October 29, 1921.

† In this paper the powers of z will be understood to be included among the polynomials.

‡ We should have been glad to use the terms "primitive" and "imprimitive" had not these terms already been applied to polynomials, with a different significance.

Two decompositions of $F(z)$ which are not equivalent will be called *distinct*, whether they contain the same number of polynomials or not.

Our task in the present paper will be, after showing the constancy of the number of prime polynomials in the distinct decompositions of a given polynomial into prime polynomials, to determine those polynomials which have two or more distinct decompositions into prime polynomials.

The results will be easy to understand after we have examined some special cases.

If $F(z)$ is a power of z with an exponent which is not a power of a prime, then to every permutation of the prime factors of the exponent there corresponds a separate decomposition of $F(z)$ into prime powers of z. Here is a situation more general in certain respects. Let

$$\varphi(z) = z^n, \qquad \alpha(z) = z^r\, g(z^n)$$
$$\psi(z) = z^r[g(z)]^n, \qquad \beta(z) = z^n,$$

where $g(z)$ is any polynomial in z. If the degrees of $\varphi(z)$ and $\alpha(z)$ are prime, and unequal to each other, the four polynomials will be prime* and we will have in the equation

$$\varphi\alpha = \psi\beta$$

two distinct decompositions of the same polynomial.

Again, it is well known that for every positive integer n, we have

$$\cos nu = f_n(\cos u),$$

where $f_n(z)$ is the "trigonometric polynomial" of degree n. It is clear that for every pair of integers m and n

$$f_{mn} = f_m f_n = f_n f_m.$$

If, now, in a decomposition of $F(z)$ into prime polynomials

$$F = \varphi_1 \varphi_2 \cdots \varphi_r, \tag{3}$$

we have for an adjacent pair of prime polynomials,

$$\varphi_i = \lambda_1 \pi_1 \lambda_2, \qquad \varphi_{i+1} = \lambda_2^{-1} \pi_2 \lambda_3,$$

where $\lambda_1(z)$, $\lambda_2(z)$ and $\lambda_3(z)$ are linear and where $\pi_1(z)$ and $\pi_2(z)$, of unequal degrees m and n, respectively, are of any of the following three types:†

* It is possible for $\alpha(z)$ to be prime even if its degree is composite. For instance, let n be any prime number, and p and q two prime numbers, so large that $pq > (p+q+2)n$. Let $pq = kn+r$, where $0 < r < n$. The most general polynomial $\alpha(z)$, of degree pq and of the form $z^r g(z^n)$ contains $k+1$ parameters. But the most general composite polynomial of degree pq contains only $p+q+2$ parameters. Hence $\alpha(z)$ is generally prime.

† Case (a) with $m = 2$ can be reduced to Case (b) by a linear transformation.

$$(a) \qquad \pi_1(z) = f_m(z), \qquad \pi_2(z) = f_n(z),$$
$$(b) \qquad \pi_1(z) = z^m, \qquad \pi_2(z) = z^r\, g(z^m),$$
$$(c) \qquad \pi_1(z) = z^r[g(z)]^n, \qquad \pi_2(z) = z^n,$$

it is clear that we have for $F(z)$ a decomposition distinct from (3),

$$F = \varphi_1 \cdots \varphi_{i-1}\psi_1\psi_2\varphi_{i+2} \cdots \varphi_r,$$

in which $\psi_1(z)$ is of degree n and $\psi_2(z)$ of degree m.

It will be shown that *if $F(z)$ has two distinct decompositions into prime polynomials, we can pass from either to a decomposition equivalent to the other by repeated steps of the three types just indicated.* This gives the solution of our problem.

The analogous problem for fractional rational functions is much more difficult. There is a much greater variety of possibilities, as one sees, without going far, on considering the formulas for the transformation of the periods of the elliptic functions. There are even cases in which the number of prime functions in one decomposition is different from that in another. We shall return to this matter in a later communication.

II. Condition for decomposability

It will give us no extra work to deal in this section with a general rational function, integral or fractional, instead of with a polynomial.*

Let w be a rational function, $F(z)$, of z, so that $z = F^{-1}(w)$. We shall understand the term "group of $F^{-1}(w)$" to mean always the group of monodromy of $F^{-1}(w)$.

We prove the following result:

A necessary and sufficient condition that $F(z)$ be composite is that the group of $F^{-1}(w)$ be imprimitive.†

Consider first the necessity of the condition. Suppose that

$$w = \varphi[\alpha(z)]$$

where $\varphi(u)$ and $\alpha(z)$ are rational functions of degrees m, greater than one, and n, greater than one, respectively. Let w_0 be any point which is not a critical point of $F^{-1}(w)$. Take a small neighborhood of w_0. Consider, in this neighborhood, any branch u_1 of $\varphi^{-1}(w)$. There are precisely n functions of w,

$$(4) \qquad z_1, z_2, \ldots, z_n,$$

* What the terms "prime," "composite," etc., will mean for a general rational function need not be explained.

† This is a more precise result, as far as the sufficiency of the condition goes, for the case with which we are dealing, than the theorem that an equation with an imprimitive Galois group can be made to depend on two equations of lower degrees.

each uniform in the above neighborhood, such that $\alpha(z_i)=u_1$ $(i=1, 2, \ldots, n)$. The relation $\alpha(z_i)=\alpha(z_1)$ holding for any i, in the above neighborhood, and continuing to hold after any substitution of the group of $F^{-1}(w)$ is applied to it, it is clear that any substitution of the group which replaces z_1 by one of the functions (4) replaces every z_i by such a function. It follows from a well known theorem that the letters (4) constitute a system of imprimitivity of the group of $F^{-1}(w)$.*

We shall say that this system of imprimitivity *is determined by* $\alpha(z)$.

Now we prove the sufficiency of the condition. The function $F^{-1}(w)$, as the inverse of a rational function, has only one pole on its Riemann surface. We shall assume that the value of w which corresponds to this pole is not a critical point of $F^{-1}(w)$. This can always be brought about by replacing $F^{-1}(w)$ by a suitable linear fractional function of itself. The inverse of this new function will be prime or composite together with $F(z)$.

Let z_1 be the branch of $F^{-1}(w)$ which has the pole, and let (4) represent a system of imprimitivity of the group of $F^{-1}(w)$. Consider the function

$$u_1 = z_1 + z_2 + \cdots + z_n.$$

This function is unchanged by those substitutions which interchange the letters of (4) among themselves. As it has a pole, it is changed by any substitution which changes the set (4) into another set. Consequently, if there are m sets such as (4), that is, if $F(z)$ is of degree mn, u_1 is uniform on a Riemann surface of m sheets. Since it has only one pole, it must be the inverse of a rational function of degree m. We will write $w=\varphi(u_1)$. Since u_1 is unchanged by those substitutions which leave z_1 fixed, it is a rational function of z_1 and w; therefore, since $w=F(z_1)$, u_1 is a rational function of z_1 alone. Let $u_1=\alpha(z_1)$. We have $w=F(z_1)=\varphi[\alpha(z_1)]$, and therefore, by the principle of the permanence of functional equations we have identically

$$F(z) = \varphi[\alpha(z)]. \tag{5}$$

In this decomposition of $F(z)$, since $\varphi(u)$ is of degree m, $\alpha(z)$ must be of degree n.

Taking now the case where $F(z)$ is a polynomial, we shall show that there exists a linear function $\lambda(z)$ such that $\varphi\lambda$ and $\lambda^{-1}\alpha$ are polynomials. Thus we would gain no generality, in our study of polynomials, by admitting fractional functions into the decompositions.

Let $\alpha(\infty)=a$. Since $\varphi\alpha$ assumes the value infinity mn times for $z=\infty$, since $\alpha(z)$ can not assume the value a more than n times for $z=\infty$, and since $\varphi(u)$ cannot assume the value infinity more than m times for $u=a$, the values a and ∞ must be assumed precisely n and m times, respectively, by $\alpha(z)$ and

* Netto, *Gruppen- und Substitutionentheorie*, Leipzig, 1908, p. 143.

$\varphi(z)$ at the indicated points. Let $\lambda(z)$ be any linear function such that $\lambda^{-1}(a) = \infty$. Then $\lambda^{-1}\alpha$ assumes the value infinity n times for $z = \infty$, and is therefore a polynomial of degree n. Similarly $\varphi\lambda$ is a polynomial of degree m.

We shall show now that any decomposition of $F(z)$ which is based on the given system of imprimitivity is equivalent to (5). Suppose that $w = \varphi_1(v)$ and $v = \alpha_1(z)$ are two rational functions of degrees m and n, respectively, such that

$$F(z) = \varphi_1[\alpha_1(z)], \tag{6}$$

where, for the functions of (4),

$$\alpha_1(z_1) = \alpha_1(z_2) = \cdots = \alpha_1(z_n). \tag{7}$$

Let $v_1 = \alpha_1(z_1)$. By (7), v_1 is unchanged by those substitutions which interchange the functions (4) among themselves. It is therefore a rational function of u_1 and w, and as $w = \varphi(u_1)$, it is a rational function of u_1 alone. Thus $\alpha_1(z)$ is a rational function of $\alpha(z)$, and as it is of the same degree as $\alpha(z)$, it must be a linear function of $\alpha(z)$. This settles the equivalence of (5) and (6).

If, in (5), $\alpha(z)$ is a prime function, the systems of imprimitivity determined by $\alpha(z)$ cannot be broken up into smaller systems of imprimitivity. For, were that possible, the sum of the functions of that smaller set which contained z_1 would be a rational function of z_1, while u_1 would be a rational function of the sum, so that $\alpha(z)$ would be composite.

As the group of $F^{-1}(w)$ has only a finite number of systems of imprimitivity, there are only a finite number of possibilities for $\varphi_r(z)$ in (2). It follows by a quick induction that $F(z)$ *has only a finite number of distinct decompositions.*

III. The invariant integers*

Suppose that a polynomial $F(z)$ has two *distinct* decompositions into prime polynomials

$$F = \varphi_1\varphi_2 \cdots \varphi_r, \qquad F = \psi_1\psi_2 \cdots \psi_s. \tag{8}$$

We propose to show that r *equals* s, and that *the degrees of the polynomials* ψ *are the same as the degrees of the polynomials* φ *except for the order in which they occur.*

The theorem just stated is certainly true if the degree of $F(z)$ does not exceed six. It will suffice therefore to show that it holds for polynomials of degree n if it holds for the polynomials of all lower degrees.

If $\varphi_r(z)$ and $\psi_s(z)$ determine the same systems of imprimitivity, they are, as was

* The group-theoretic part of the present section consists mainly of a proof that Jordan's incorrect result on the invariance of the factors of imprimitivity of a group, *Traité des Substitutions*, p. 34, and Giornale di Matematiche, vol. 10 (1872), p. 176, holds when the group contains a substitution which permutes all of the letters in a single cycle.

shown in the preceding section, linear functions of each other. Making the substitution $\varphi(z)=u$ in both representations of $F(z)$, we are led to the consideration of a polynomial of degree less than n. The induction is thus carried through for this special case.

To handle the case where $\varphi_r(z)$ and $\psi_s(z)$ determine distinct types of imprimitivity, it will be desirable to determine the possible systems of imprimitivity of the group of $F^{-1}(w)$. This is made possible, fortunately, by the fact that $F^{-1}(w)$, if $F(z)$ is a polynomial of degree n, has a branch point of order $n-1$ at infinity. If the branches of $F^{-1}(w)$ are suitably numbered the substitution corresponding to this branch point may be written in a single cycle

$$(12 \cdots n). \tag{9}$$

Every system of imprimitivity of the group of $F^{-1}(w)$ is respected by this substitution. Suppose that there is a type of imprimitivity in which the letters break up into p sets, with h letters in each set, so that $n=hp$. Let the elements of the set which contains 1, arranged in order of magnitude, be

$$1, a_1, a_2, \ldots, a_{h-1}. \tag{10}$$

The (a_1-1)th power of (9), since it replaces 1 by a_1, interchanges the letters of (10) among themselves. It cannot disturb the cyclic order of the letters. Hence it must replace each letter by the one which follows it cyclically. We must therefore have $a_1=p+1$, and the set (10) is

$$1, \quad p+1, \quad 2p+1, \quad \cdots, \quad (h-1)\,p+1. \tag{11}$$

If we operate on the elements of (11) with the $(\nu-1)$th power of (9) we obtain the set $\nu+ip$ $(i=1, 2\ldots, h-1)$, which is also a system of imprimitivity.

Thus, given any divisor h, of n, the group of $F^{-1}(w)$ can have at most one type of imprimitivity with h letters in each set. Hence, if, in (8), $\varphi_r(z)$ and $\psi_s(z)$ determine distinct types of imprimitivity, these polynomials must be of different degrees.

Suppose then, that $\varphi_r(z)$ is of degree h, with $n=hp$, and that $\psi_s(z)$ is of degree k, with $n=kq$. The type of imprimitivity determined by $\varphi_r(z)$ has the sets

$$\nu+ip \qquad \begin{pmatrix} \nu=1, 2, \ldots, p \\ i=0, 1, \ldots, h-1 \end{pmatrix}, \tag{12}$$

each value of ν determining one set, and the type of imprimitivity determined by $\psi_s(z)$ has the sets

$$\nu+iq \qquad \begin{pmatrix} \nu=1, 2, \ldots, q \\ i=0, 1, \ldots, k-1 \end{pmatrix}. \tag{13}$$

We shall prove now a fact which will be of the greatest importance in the following section, namely that *h and k are relatively prime*. Suppose that h and k had a factor δ in common. The substitution (9) and its powers have a type of imprimitivity with δ letters in each set, each set of which is contained in a set of (12) and in a set of (13). Hence by the next to the last paragraph of §II, the existence of the factor δ would lead to the conclusion that $\varphi_r(z)$ and $\psi_s(z)$ are not prime.

As n is divisible by hk, the substitution (9) and its powers have systems of imprimitivity containing hk letters. Let I denote that one of these systems which contains z_1. Then I can be considered as consisting either of k sets of (12) or of h sets of (13). As no set of (12) has more than one letter in common with a set of (13), each of the k sets of (12) has precisely one letter in common with each of the h sets of (13).

Consider those substitutions of the group of $F^{-1}(w)$ which either leave z_1 fixed or replace it by some other letter of I. If we can show that these substitutions interchange the letters of I among themselves, we will know that *I is a system of imprimitivity of the group of $F^{-1}(w)$*. Such a substitution replaces the set of (12) to which z_1 belongs by another set of (12). Since both of these sets of (12) contain elements of each of the h sets of (13) which make up I, the substitution must interchange the h sets of (13), and hence the letters of I, among themselves.

There is a polynomial $\mu(z)$, of degree hk, which determines the system of imprimitivity just shown to exist. We have $F=\sigma\mu$ where $\sigma(z)$ is some polynomial, and also

$$\mu = \zeta\varphi_r, \qquad \mu = \xi\psi_s$$

where $\zeta(z)$ and $\xi(z)$ are polynomials of degrees k and h respectively.

The polynomial $\zeta(z)$ must be prime. If not, the set I would break up into several new systems of imprimitivity, each of these systems containing more than one set of (12). Such a new system would have more than one letter, but less than k letters, in common with each of the h sets of (13) of which I is composed. Thus the sets of (13) would break up into smaller systems of imprimitivity, and $\psi_s(z)$ could not be prime. Similarly, $\xi(z)$ must be prime.

Consider the two decompositions

$$F = \varphi_1\varphi_2 \cdots \varphi_r, \qquad F = \sigma\zeta\varphi_r. \tag{14}$$

We have

$$\varphi_1\varphi_2 \cdots \varphi_{r-1} = \sigma\zeta,$$

and since the members of this last equation are of degree less than n, the second member, with $\sigma(z)$ decomposed into prime polynomials, would contain the same

number of polynomials as the first, with the same degrees. Similar remarks apply to the two decompositions

$$F = \psi_1\psi_2 \cdots \psi_s, \qquad F = \sigma\xi\psi_s. \tag{15}$$

We need only to compare the situations described in (14) and (15) to complete the induction necessary for the proof of our theorem.

We shall now make clear by what kind of steps it is possible to pass from one of the decompositions of (8) to the other. We say that there is a sequence of decompositions of $F(z)$, beginning with the first of (8) and ending with the second, such that any decomposition is either equivalent to the one which precedes it, or else can be formed from the one which precedes it by taking two adjacent prime polynomials, $\varphi(z)$ and $\alpha(z)$ of the latter, and replacing $\varphi\alpha$ by $\psi\beta$, where $\psi(z)$ has the same degree as $\alpha(z)$ and $\beta(z)$ the same degree as $\varphi(z)$. That is, two consecutive decompositions, if not equivalent, have the forms

$$F = \zeta_1 \cdots \zeta_i\varphi\alpha\zeta_{i+3} \cdots \zeta_r$$

and

$$F = \zeta_1 \cdots \zeta_i\psi\beta\zeta_{i+3} \cdots \zeta_r$$

respectively.

This result is readily seen to hold if the degree of $F(z)$ does not exceed six. We shall prove that it holds for polynomials of degree n if it holds for polynomials of all lower degrees.

If ϕ_r and ψ_r, in (8), determine the same systems of imprimitivity, they are linear functions of each other. Let $\psi_r = \lambda\varphi_r$, and put $\psi'_{r-1} = \psi_{r-1}\lambda^{-1}$. It is possible to pass from the first of the two decompositions

$$\varphi_1\varphi_2 \cdots \varphi_{r-1}\,\varphi_r, \qquad \psi_1\psi_2 \cdots \psi_{r-2}\psi'_{r-1}\varphi_r$$

to the second by steps of the types described above, since

$$\varphi_1\varphi_2 \cdots \varphi_{r-1}$$

is of degree less than n. The second of these two decompositions is equivalent to the second of (8).

If φ_r and ψ_r determine distinct systems of imprimitivity we can pass from the first decomposition of (8) to the decomposition $\sigma\zeta\varphi_r$ by steps of the above types,* since $\sigma\zeta$ is of degree less than n. We can pass from the second decomposition of (8) to the decomposition $\sigma\xi\psi_r$ by similar steps. It needs only one step to pass from $\sigma\zeta\varphi_r$ to $\sigma\xi\psi_r$.

Consider any three consecutive decompositions of the sequence. To fix our ideas, suppose that the second is equivalent to the first, and that the third results

* We understand of course that σ is decomposed into prime polynomials.

from the second by a change of $\varphi\alpha$ into $\psi\beta$ as above described. It is clear that we could suppress the second decomposition and pass from the first to a decomposition equivalent to the third by redecomposing a polynomial $\varphi\alpha$.

Continuing this suppression of equivalent decompositions, we see that there exists a sequence of decompositions, beginning with the first of (8), and ending with one equivalent to the second, each one of which is obtained from the preceding by redecomposing a polynomial $\varphi\alpha$.

IV. The two-polynomial problem

The results of the preceding section reduce our problem to the following:
Under what circumstances can we have

$$\varphi\alpha = \psi\beta,$$

*where $\varphi(z)$ and $\beta(z)$ are two prime polynomials of degree m, greater than unity, and $\psi(z)$ and $\alpha(z)$ two prime polynomials of degree n, greater than unity?**

In our treatment of this problem we shall not have occasion to impose the condition that the polynomials be prime. We shall assume only that each system of imprimitivity determined by $\alpha(z)$ has precisely one letter in common with each system of imprimitivity determined by $\beta(z)$. For this it is necessary and sufficient that m and n be relatively prime.

We put

$$w = F(z) = \varphi(u) = \psi(v),$$

where

$$u = \alpha(z), \qquad v = \beta(z).$$

There will be five Riemann surfaces to consider, those namely for $F^{-1}(w)$, $\varphi^{-1}(w)$, $\psi^{-1}(w)$, $\alpha^{-1}(u)$ and $\beta^{-1}(v)$.

It will be very important to see how the surface for $F^{-1}(w)$ is related to those for $\varphi^{-1}(w)$ and $\alpha^{-1}(u)$. Suppose that for $u=c$, $\alpha^{-1}(u)$ has a critical point with a certain number of cycles. Since $\varphi^{-1}(u)$ assumes no value more than once on its Riemann surface, $F^{-1}(w)$ will surely have a critical point for $w=\varphi(c)$. If the value c is assumed by a branch u_1 of $\varphi^{-1}(w)$ which is uniform in the neighborhood of $\varphi(c)$, those branches z of $F^{-1}(w)$ for which $\alpha(z)=u_1$ will be ramified at $\varphi(c)$ like the branches of $\alpha^{-1}(u)$ for $u=c$. If the value c is assumed by a cycle of r branches of $\varphi^{-1}(w)$, each cycle of $\alpha^{-1}(u)$ at $u=c$ will lead to a cycle of $F^{-1}(w)$ at

* An idea which presents itself naturally is to consider this problem as one in undetermined coefficients. One might hope, for instance, with a judicious use of linear transformations, to show without actually determining the coefficients of the polynomials, that aside from Cases (*b*) and (*c*) mentioned in the introduction there is only one possibility, which would, of course, have to be Case (*a*). A study of the equations for the coefficients convinces me that such a plan would not be easy to carry out, and that the function-theoretic methods used here are not far-fetched.

$\varphi(c)$, with r times as many sheets. If $\varphi^{-1}(w)$ has a critical point for $w=d$, and if $\alpha^{-1}(u)$ has no critical point for any value of $\varphi^{-1}(d)$, then each cycle of $\varphi^{-1}(w)$ at d leads to n cycles of the same number of sheets for $F^{-1}(w)$ at d.

In any case, if $\varphi^{-1}(w)$ has no uniform branches at d, $F^{-1}(w)$ will have no uniform branches at d.

If the branches of $\varphi^{-1}(w)$ are $u_1, u_2, \ldots, u_m$, the group of $F^{-1}(w)$ will have m systems of imprimitivity

$$U_1, U_2, \ldots, U_m,$$

with n letters z in each set, such that for every letter of U_i we have $\alpha(z)=u_i$. Similarly, if the branches of $\psi^{-1}(w)$ are $v_1, v_2, \ldots, v_n$, the group of $F^{-1}(w)$ will have n systems of imprimitivity

$$V_1, V_2, \ldots, V_n,$$

with m letters z in each set.

If w moves around a closed path the sets U are permuted like the branches of $\varphi^{-1}(w)$, and the sets V like the branches of $\psi^{-1}(w)$.

The following fact will be fundamental in our work:

*If a substitution of the group of $F^{-1}(w)$ interchanges the letters of some set U_i among themselves, it interchanges the sets V with a substitution similar to that which it effects on the letters of U_i.**

This follows from the fact that U_i contains precisely one letter of each set V.

It will be convenient in what follows to call the sum of the orders of the branch points of an algebraic function at a given critical point the *index* of the function at the point. Since the Riemann surface of the inverse of a rational function is of genus zero, the sum of the indices of the inverse of a rational function of degree m, for all of its critical points, is $2m-2$. The sum of the indices of the inverse of a polynomial of degree m, for all of its critical points excluding infinity, is $m-1$. If the inverse of a rational function of degree m has a critical point for which none of its branches is uniform, its index at that point is at least $m/2$.

We are going to derive some relations between the critical points of $\alpha^{-1}(u)$ and those of $\psi^{-1}(w)$. Similar results will hold for $\beta^{-1}(v)$ and $\varphi^{-1}(w)$.

Consider a critical point $u=c$ of $\alpha^{-1}(u)$. As u makes a turn around c the branches of $\alpha^{-1}(u)$ undergo a certain substitution. Suppose that $\varphi^{-1}(w)$ has a branch u_i which is uniform in the neighborhood of $\varphi(c)$ and assumes there the value c. As w makes a turn about $\varphi(c)$ the value of u_i makes a turn about c. The letters of the set U_i will undergo a permutation similar to that of the branches of $\alpha^{-1}(u)$ for the point c. The sets V must consequently undergo a similar permutation.

* In saying that two substitutions are similar, we will mean that they consist of the same number of cycles, with an equal number of letters in corresponding cycles.

Hence, in this case, $\psi^{-1}(w)$ has a critical point for $w=\varphi(c)$, with a substitution similar to that of $\alpha^{-1}(u)$ for $u=c$. The index of $\psi^{-1}(w)$ at $\varphi(c)$ equals the index of $\alpha^{-1}(u)$ at c.

If a value d is assumed at $\varphi(c)$ by another branch of $\varphi^{-1}(w)$, also uniform at $\varphi(c)$, then $\alpha^{-1}(u)$ must have a critical point at d, with a substitution similar to that which it has at c. In this case the index of $\psi^{-1}(w)$ at $\varphi(c)$ is less than the sum of the indices of $\alpha^{-1}(u)$ for c and d.

Let us see now conversely, and with a little greater precision, what the nature of a critical point of $\psi^{-1}(w)$ implies with respect to the critical points of $\alpha^{-1}(u)$.

Suppose that at the point $w=e$, $\varphi^{-1}(w)$ has a critical point with the substitution

$$S_1 S_2 \ \dots \ S_r, \tag{16}$$

where each S_ρ represents a cycle containing i_ρ letters, and $\psi^{-1}(w)$ a critical point with the substitution

$$T_1 T_2 \ \dots \ T_s, \tag{17}$$

each T_ρ representing a cycle containing j_ρ letters.

Consider a cycle S_ρ, and let h be the value which $\varphi^{-1}(w)$ assumes at the branch point corresponding to that cycle. We seek to determine how the branches of $\alpha^{-1}(u)$ behave as u makes a turn around h. For u to turn once around h, w must turn i_ρ times around e. For this turn, the branches of $\psi^{-1}(w)$ undergo a permutation which is the i_ρth power of (17). The function $\alpha^{-1}(u)$ has a critical point for $u=h$ with a substitution similar to the i_ρth power of (17).

Let us get an idea of the index of $\alpha^{-1}(u)$ at h. Consider one of the cycles T_q. If i_ρ is prime to j_q, the i_ρth power of T_q will be a cyclic substitution of order j_q, so that the cycle T_q will contribute j_q-1 to the index of $\alpha^{-1}(u)$ at h. If i_ρ is divisible by j_q, T_q contributes nothing to the index of $\alpha^{-1}(u)$ at h. If i_ρ is not divisible by j_q, then at the worst the i_ρth power of T_q can break up into cycles of two letters, so that T_q must contribute at least $j_q/2$ to the index of $\alpha^{-1}(u)$ at h.

If the sum of the indices of $\alpha^{-1}(u)$ for those points which correspond to the values assumed by $\varphi^{-1}(w)$ at e is less than the index of $\psi^{-1}(w)$ at e we shall say that $\psi^{-1}(w)$ has an *extra point* at e. This convention will apply also with respect to $\beta^{-1}(v)$ and $\varphi^{-1}(w)$.

Such an extra point of $\psi^{-1}(w)$ must exist, for instance, if $\psi^{-1}(w)$ has a critical point at which more than one branch of $\varphi^{-1}(w)$ is uniform, since, $\alpha(z)$ and $\psi(v)$ having the same degree, the sum of the indices of the inverse of one equals the corresponding sum for the other. If $\psi^{-1}(w)$ has an extra point at e, no branch of $\varphi^{-1}(w)$ can be uniform in the neighborhood of e, for, as seen above, every such uniform branch would lead to a critical point of $\alpha^{-1}(u)$ with an index equal to that of $\psi^{-1}(w)$ at e. Also $\psi^{-1}(w)$ cannot have an extra point at infinity, since

$\psi^{-1}(w)$ and $\alpha^{-1}(u)$ have the same index, $n-1$, at infinity. Finally, $\psi^{-1}(u)$ cannot have more than a single extra point. In short, at each extra point of $\psi^{-1}(w)$, the function $\varphi^{-1}(w)$, having no uniform branches, has an index not less than $m/2$, so that, since the sum of the indices of $\varphi^{-1}(w)$ for all of its critical points, excluding infinity, is $m-1$, two such extra points cannot exist.

We are going to show that $\psi^{-1}(w)$ and $\varphi^{-1}(w)$ cannot each have an extra point.

If both functions have an extra point, their extra points must correspond to the same value of w. For at an extra point either of $\psi^{-1}(w)$ or of $\varphi^{-1}(w)$, the function $F^{-1}(w)$, since every one of its branches is permuted, must have an index not less than $mn/2$. There cannot be two such points.

Suppose then that $\psi^{-1}(w)$ and $\varphi^{-1}(w)$ each have an extra point at $w=e$. Neither can have a branch which is uniform at e. The branches of $\varphi^{-1}(w)$ undergo at e a substitution of the form (16) with

$$i_1 + i_2 + \cdots + i_r = m, \tag{18}$$

while the branches of $\psi^{-1}(w)$ undergo a substitution of the form (17), with

$$j_1 + j_2 + \cdots + j_s = n. \tag{19}$$

Suppose that every one of the numbers j has one of the following two properties:*

(a) Two of the numbers i are not divisible by it.

(b) One of the numbers i has no factor in common with it.

Consider j_q, for instance, and suppose that it has the property (a) relative to i_a and i_b. Let h and k be the values which $\varphi^{-1}(w)$ assumes at the branch points corresponding to the cycles S_a and S_b. As i_a is not divisible by j_q, the cycle T_q must contribute at least $j_q/2$ to the index of $\alpha^{-1}(u)$ at h. Similarly it must contribute at least $j_q/2$ to the index of $\alpha^{-1}(u)$ at k. Again, if j_q has the property (b) relative to i_a, T_q contributes precisely j_q-1 to the index of $\alpha^{-1}(u)$ at h.

It is clear that if every j had one of the two properties, the sum of the indices of $\alpha^{-1}(u)$ at the points corresponding to the values which $\varphi^{-1}(w)$ assumes at e would be at least as great as the index of $\psi^{-1}(w)$ at e, so that e could not be an extra point of $\psi^{-1}(w)$.

There must consequently exist a number j_q which divides into all except perhaps one of the numbers i, and which has a factor in common with that i. That factor of j_q, being a factor of all of the numbers i, is a factor of m, and since m and n are prime to each other it cannot be a factor of n. Thus there is a j which is not divisible by that factor of the numbers i. Let it be j_1, for instance, and let h be the value which $\psi^{-1}(w)$ assumes at the branch point corresponding to the cycle T_1. Since j_1 is not divisible by any i_p, each cycle S_p contributes at least $i_p/2$ to the index of $\beta^{-1}(v)$ at h.

* If the substitution (16) has only one cycle, only the second property need be considered.

Hence the index of $\beta^{-1}(v)$ at h is at least $m/2$. Certainly then, the sum of the indices of $\beta^{-1}(v)$ for the points corresponding to the values assumed by $\psi^{-1}(w)$ at e is not less than $m/2$.

By an argument similar to that employed above, we can show that the sum of the indices of $\alpha^{-1}(u)$ at the points corresponding to the values assumed by $\varphi^{-1}(w)$ at e is at least $n/2$.

Let the indices of $\varphi^{-1}(w)$ and $\psi^{-1}(v)$ at e be

$$\frac{m}{2} + x, \qquad \frac{n}{2} + y,$$

respectively. The numbers x and y, each of which is at least unity, need not be integral.

Suppose, to fix our ideas, that x is not less than y. Since the sums of the indices of $\alpha^{-1}(u)$ and of $\psi^{-1}(w)$ are the same, namely $2n-2$, $\alpha^{-1}(u)$ must have a critical point with a finite affix which does not correspond to a value of $\varphi^{-1}(w)$ at e. Let c be such a critical point of $\alpha^{-1}(u)$. Consider the behavior of $\varphi^{-1}(w)$ at $\varphi(c)$. Since the sum of the indices of $\varphi^{-1}(w)$, excluding the point ∞, is $m-1$, the index of $\varphi^{-1}(w)$ at $\varphi(c)$ is not greater than

$$m - 1 - \left(\frac{m}{2} + x\right) = \frac{m}{2} - x - 1.$$

Hence not more than $m-2x-2$ branches of $\varphi^{-1}(w)$ can be permuted at $\varphi(c)$. Then at least $2x + 2$ branches of $\varphi^{-1}(w)$ are uniform in the neighborhood of $\varphi(c)$. Consequently $\psi^{-1}(w)$ must have a branch point at $\varphi(c)$ with an index at least that of $\alpha^{-1}(u)$ at c, and $\alpha^{-1}(u)$ must have at least $2x + 2$ such points, including c. The index of $\psi^{-1}(w)$ at $\varphi(c)$ is at least unity. Hence the sum of the indices of $\alpha^{-1}(u)$ at the $2x + 2$ or more critical points under consideration exceeds the index of $\psi^{-1}(w)$ at $\varphi(c)$ by at least $2x + 1$, a quantity greater than y. Thus the extra point at e fails to give $\varphi^{-1}(v)$ a sufficiently large index to make up for the situation at the point $\varphi(c)$.

The proof that $\psi^{-1}(w)$ and $\varphi^{-1}(w)$ cannot each have an extra point is complete. It will not be difficult now to settle the two-polynomial problem.

Suppose that it is $\psi^{-1}(w)$ which has no extra point. Then at any critical point of $\psi^{-1}(w)$, $\varphi^{-1}(w)$ cannot have more than one uniform branch. It follows that $\psi^{-1}(w)$ cannot have more than two critical points in addition to infinity.

Suppose first that $\psi^{-1}(w)$ has only one critical point in addition to infinity. At this point, $\psi^{-1}(w)$ must have a branch point of order $n-1$. We see that $\psi(v)$ is of the form $a\,(v + b)^n + c$. At this critical point of $\psi^{-1}(w)$, $\varphi^{-1}(w)$ has at most one uniform branch.

Suppose that for this point the branches of $\varphi^{-1}(w)$ undergo the substitution (16), where, understanding that there may be one cycle which consists of a single

letter, equation (18) holds. If there were two of the numbers i in (18) which were not divisible by n, $\alpha^{-1}(u)$ would have two critical points, distinct from infinity, each with an index not less than $n/2$. Therefore all but one of the numbers i must be divisible by n, so that one i must be prime to n.* It follows that $\alpha^{-1}(u)$ has a branch point of order $n-1$, its only critical point in addition to infinity, and that $\alpha(z)$ is of the form $d\ (z+e)^n + f$.

If w, u, v and z are subjected to appropriate linear transformations, the case under consideration can be reduced to that in which $\psi(v)=v^n$ and $\alpha(z)=z^n$. We have thus to determine under what circumstances it is possible to have

$$\varphi(z^n) = [\beta(z)]^n.$$

We must have, in this case, denoting by ϵ any primitive nth root of unity,

$$\beta(\epsilon z) = \epsilon^r \beta(z)$$

where r is some positive integer not greater than n. Consequently

$$\beta(z) = z^r\, g(z^n),$$

where $g(z^n)$ is a polynomial in z^n. Hence

$$\varphi(u) = u^r [g(u)]^n.$$

Let us summarize these results for the case where $\psi^{-1}(w)$ has only one critical point in addition to infinity. We must have

$$\varphi = \lambda_1 \varphi_1 \lambda_2, \qquad \alpha = \lambda_2^{-1} \alpha_1 \lambda_3,$$
$$\psi = \lambda_1 \psi_1 \lambda_4, \qquad \beta = \lambda_4^{-1} \beta_1 \lambda_3,$$

where $\lambda_1(z)$, $\lambda_2(z)$, $\lambda_3(z)$ and $\lambda_4(z)$ are linear, and where

$$\varphi_1(u) = u^r [g(u)]^n, \qquad \alpha_1(z) = z^n,$$
$$\psi_1(v) = v^n, \qquad \beta_1(z) = z^r [g(z^n)],$$

where r is any positive integer prime to n, and where $g(u)$ is any polynomial in u. If $g(u)=1$, we have the simple case of two power functions.†

Consider now the case where $\psi^{-1}(w)$ has two critical points besides infinity. The index of $\varphi^{-1}(w)$ at each of these points is at least $(m-1)/2$. It follows that the index of $\varphi^{-1}(w)$ at each of these points is precisely $(m-1)/2$, and that at each point $\varphi^{-1}(w)$ has one uniform branch and $(m-1)/2$ branch points of the first order. Of course m must be odd.

At each of the finite critical points of $\psi^{-1}(w)$, the order of the substitution which its branches undergo must be two. Otherwise, since $\varphi^{-1}(w)$ would have, at such a critical point, cycles with a number of sheets not divisible by the order

* If (18) consists of one cycle this result follows immediately.

† Except in the case of $r=m$, $\varphi^{-1}(w)$ will have an extra point.

of the substitution just mentioned, and also a uniform branch, $\psi^{-1}(w)$ would necessarily have an extra point. Hence the index of $\psi^{-1}(w)$ at either of its two finite critical points cannot exceed $n/2$. Hence if n is even, $\psi^{-1}(w)$ has one critical point with $n/2$ branch points, and one critical point with two uniform branches and $n/2-1$ branch points of the first order.* If n is odd, $\psi^{-1}(w)$ must have, at each of its finite critical points, one uniform branch and $(n-1)/2$ branch points of the first order.

Let us take the case in which n is even. At the points which correspond to the values assumed at the critical points of $\varphi^{-1}(w)$ by its uniform branches, $\alpha^{-1}(u)$ has critical points of the types of those of $\psi^{-1}(w)$. Thus $F^{-1}(w)$, since $F=\varphi\alpha$, has one critical point at which all of its branch points are permuted in pairs, and one at which two branches are uniform and the others are permuted in pairs.

We may assume that the branches of $F^{-1}(w)$ are so numbered that the cyclic substitution at infinity, which we shall denote by s_0, is given by the formula

$$\nu' \equiv \nu + 1 \pmod{mn}.$$

Let s_1 represent the substitution at the finite critical point where no branches are uniform and s_2 the substitution at the second finite critical point. We have $s_0s_1=s_2$. Suppose that s_1 replaces i by j. Then s_0s_1 replaces $i-1$ by j. But since s_2 is of order two, s_0s_1 must replace j by $i-1$; that is, s_1 must replace $j+1$ by $i-1$. Similarly s_1 must replace $j+2$ by $i-2$, etc. It is clear then that s_1 may be written in the form.

$$\nu' \equiv -\nu + k \pmod{mn},$$

where, since s_1 leaves no branch fixed, k must be odd. If we renumber the branches of $F^{-1}(w)$, giving to the branch numbered ν the number $\mu=\nu-(k-1)/2$, the substitutions s_0 and s_1 become

$$\mu' = \mu + 1, \qquad \mu' = -\mu + 1,$$

respectively.

Similar results are obtained if n is odd.

Thus only one mechanism is possible for the surface of $F^{-1}(w)$. The two finite critical points can be placed at pleasure by subjecting w to an appropriate linear transformation. Hence if $f_{mn}(z)$ is a particular function of degree mn which has two decompositions of the types considered, we will have

$$F = \lambda_1 f_{mn} \lambda_2$$

where λ_1 and λ_2 are linear. But the trigonometric polynomial $f_{mn}(z)$, defined by the relation

$$\cos mnu = f_{mn}(\cos u),$$

* In this case $\varphi^{-1}(w)$ has an extra point.

is precisely of this type.* We may, then, take $f_{mn}(z)$ as a trigonometric polynomial. Now $F(z)$ has the two decompositions.

$$F = \lambda_1 f_m f_n \lambda_2, \qquad F = \lambda_1 f_n f_m \lambda_2.$$

As the group of $F^{-1}(w)$ can have only one type of imprimitivity with a given number of letters in each set, and as any two decompositions based on the same system of imprimitivity are equivalent, we must have

$$\varphi = \lambda_1 f_m \lambda_3, \qquad \alpha = \lambda_3^{-1} f_n \lambda_2,$$
$$\psi = \lambda_1 f_n \lambda_4, \qquad \beta = \lambda_4^{-1} f_m \lambda_2,$$

where $\lambda_3(z)$ and $\lambda_4(z)$ are linear. This settles the case where $\psi^{-1}(w)$ has two critical points in addition to infinity.

A consideration of the material of this section and of the preceding one leads to the statement of results made in the introduction. We shall not enlarge upon that statement.

---

* It is easy to show that the inverses of the trigonometric polynomials have, in addition to infinity, the critical points 1 and -1, where their branches are permuted in pairs.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

SOME TWO-DIMENSIONAL LOCI CONNECTED WITH CROSS RATIOS*

BY

J. L. WALSH

1. Introduction. The writer has recently published a proof of the following theorem:†

THEOREM I. *If the points z_1, z_2, z_3 vary independently and have circular regions as their respective loci, then the locus of the point z_4 defined by the real constant cross ratio*

$$\lambda = (z_1, z_2, z_3, z_4)$$

is also a circular region.

It is the purpose of the present paper to consider generalizations of and other results related to Theorem I, primarily the determination of the locus of the point z_4 defined as in Theorem I, when z_1, z_2, z_3 vary independently so as to have certain prescribed loci. Thus one may raise the following question: If two variable points lie respectively on two fixed circles, where does the mid-point of their segment lie? The answer is contained in Theorem VI. Or again, if the loci of the points z_1, z_2, z_3 are regions each bounded by a number of circles, what can be said of the locus of z_4? The answer is given by Theorem VII.

All the results proved concerning such loci as these can be interpreted in terms of the roots of the jacobian of two binary forms. Such an interpretation is given in §12.

The chief method used in the present paper to determine the locus of z_4 in any given case is that suggested in I (pp. 102, 103, footnote), namely the determination of the locus of z_4 when z_1 and z_2 are held fast and z_3 varies over its locus; the determination of the locus of this locus of z_4 when z_1 is kept fixed but z_2 varies over its locus; and finally the determination of the locus of this new locus of z_4 when z_1 varies over its prescribed locus. In the present paper this method is carried through geometrically. The same method has been used by Professor A. B. Coble‡ to determine analytically the locus of z_4 in Theorem I,

* Presented to the Society, September 7, 1920, and September 9, 1921.

† These Transactions, vol. 22 (1921), pp. 101–116; this paper will be referred to as I. We shall also have occasion to refer to another of our papers, using the letter S: *Comptes Rendus du Congrès International des Mathématiciens*, Strasbourg, 1920, pp. 349–352.

The term *locus* used in Theorem I of the present paper replaces the term *envelope* used in I.

‡ Bulletin of the American Mathematical Society, vol. 27 (1920–1921), pp. 434–437.

but analytic determination in the more general case is less illuminating and probably more difficult than the geometric determination.

Throughout the present paper, as throughout I, circles play a central rôle.

2. Theorem II, a property of the boundaries of the loci. We suppose once for all that the loci of the points z_1, z_2, z_3 are closed regions bounded by a finite number of regular curves.* We admit, however, the possibility of having these loci either points or regular curves instead of two-dimensional continua bounded by such curves.

For loci of this sort we prove a result which corresponds to, but is more general than Lemma III (I, p. 104); the prescribed loci of z_1, z_2, z_3 are denoted by T_1, T_2, T_3, respectively, and the locus of z_4 defined as in Theorem I is denoted by T_4.

THEOREM II. *If the point z_4 is on the boundary of T_4, then any set of points z_1, z_2, z_3 corresponding lie on the boundaries of their respective regions T_1, T_2, T_3; if none of these four points is at a vertex of its proper region, the circle C through the points z_1, z_2, z_3, z_4 cuts the boundaries of T_1, T_2, T_3, T_4 all at angles of the same magnitude, and if C is transformed into a straight line, the lines tangent to these boundaries at the points z_1, z_2, z_3, z_4, respectively, are parallel. If one or more of the points z_1, z_2, z_3, z_4 is at a vertex of its proper region, C cuts the boundaries of the other regions at angles of the same magnitude, and if C is transformed into a straight line the lines tangent to these boundaries at the corresponding points are parallel.*

If we consider the defining relation to be

$$(z_1, z_2, z_3, z_4) = \lambda,$$

if $\lambda = 0$, 1, or ∞, and if T_4 coincides with one of the original regions (see I, p. 103), only one of the points z_1, z_2, z_3 is effectively concerned with the location of z_4, and Theorem II is not true. Whenever T_4 is the entire plane, the locus of z_4 has no boundary and the theorem has no meaning. These possibilities are henceforth excluded.

We phrase the proof to deal with the case that each of the regions T_1, T_2, T_3 is a two-dimensional continuum, but a change in wording rather than of reasoning is all that is required if one or more of these regions is a point. We discuss later the possibility of curves instead of two-dimensional continua.

When any two of the points z_1, z_2, z_3, z_4 are kept fixed, the relation defining z_4 becomes a linear relation between the other two points. Thus, when z_2 and z_3 are kept fixed, motion of z_1 over a two-dimensional continuum moves z_4 over a two-dimensional continuum, so if z_4 is on the boundary of its locus, z_1 is also on the boundary of its locus, and similarly for z_2 and z_3. Under such a linear correspondence, moreover, a circle corresponds to a circle. For example, when z_3

* For the definition of a regular curve, see Osgood, *Funktionentheorie* (second edition), pp. 51, 324, 150.

and z_4 are considered fixed and when z_1 moves along C, z_2 also moves along C. When z_1 moves from C along the boundary of T_1, z_2 moves from C and makes the same angle with C as the angle between C and the boundary of T_1. If the boundary of T_2 does not make this same angle with C, and if neither z_1 nor z_2 is at a vertex of the corresponding boundary, this motion of z_1 in one sense or the other from C will cause z_2 to move into the *interior* of T_2. Then we keep z_1 and z_3 fixed. Motion of z_3 over a two-dimensional continuum all of which is interior to T_2 will cause z_4 to move over a two-dimensional continuum all of which is interior to T_4, so z_4 cannot be on the boundary of T_4.

Thus the boundaries of T_1, T_2, T_3 all make the same angle with C, and if we keep z_1 and z_2 fixed, moving z_3 and z_4, we see by consideration of the angles at these latter two points that the boundaries of T_3 and T_4 make the same angle with C. The *direct conformality* of the angles in every case here shows that when C is transformed into a straight line the tangents to the boundaries of T_1, T_2, T_3, T_4 at z_1, z_2, z_3, z_4 are parallel if these points are not at vertices of those boundaries. If one or more of the points z_1, z_2, z_3, z_4 is at a vertex of the corresponding boundary, the result always holds for such of those points as are not at vertices of those boundaries.

If one of the original regions T_1, T_2, T_3 is a curve instead of a point or a two-dimensional continuum, further proof is necessary. If for example T_1 and T_2 are curves and if C does not cut them at z_1 and z_2 at equal angles, let us transform z_3 to infinity. Denote by C'_2 the curve of all points z_4 which correspond to a particular point z_1 of T_1 and to z_3, as z_2 traces T_2. The angle between C and C'_2 at z_4 is the same as the angle between C and C_2 at z_2. When z_1 now moves on T_1, points of C'_2 which correspond to particular points z_2 of T_2 move in the direction of motion of z_1. If C does not cut T_1 and T_2 at the same angle, this motion of z_1 on T_1 will therefore cause C'_2 to sweep out an entire two-dimensional continuum in the neighborhood of z_4 and hence z_4 cannot be on the boundary of T_4.

To compare the angles in which C cuts T_1 and the boundary of T_4, we hold z_2 and z_3 fast. Motion of z_1 from C along T_1 causes z_4 to move from C along a curve through z_4 either tangent to C or extending on both sides of C. But all points of this curve are points of T_4, and the curve cuts C at the same angle as T_1. Hence either T_4 has a vertex at z_4 or its boundary cuts C at the same angle as does T_1. We have now given all the essential reasoning in the proof of Theorem II.

We remark incidentally that curves related as are the curves bounding these regions T_1, T_2, T_3, T_4 have interesting geometric properties.* Thus when the

* In three-dimensional space we may consider the analogous problem of finding the locus of points z_4 determined by their real constant cross ratio with the points z_1, z_2, z_3 whose loci are respectively either three surfaces S_1, S_2, S_3 or three regions bounded by these three surfaces; the term *cross ratio* is to be interpreted precisely as in the plane—the four points z_1, z_2, z_3, z_4 are to be concyclic and the cross ratio can be defined by any complex coördinate system in

region T_3 is simply the point at infinity, the bounding curves C_1, C_2, C_4 are so related that C (in this case a line) sets up a correspondence between those curves, so that tangents to C_1 and C_2 at corresponding points are parallel and the corresponding points z_1, z_2, z_4 on the line C divide C in a constant ratio. Whenever three curves C_1, C_2, C_4 have this property, it is true that lines tangent to these curves at corresponding points z_1, z_2, z_4 are parallel, the centers of curvature of C_1, C_2, C_4 at z_1, z_2, z_4 are collinear and the ratios of the segments cut on the line joining them equal the ratios of the corresponding segments of the line C. If the curves C_1 and C_2 are parallel, C_4 is parallel to them.

3. Theorems III and IV; cases where the loci are curves. There are two particularly interesting types of curves T_1, T_2, T_3 which may be chosen so that T_4 shall also be a curve.

THEOREM III. *Let C_1 and C_2 be two curves which are the loci of the points z_1 and z_2 respectively; denote by C the locus of the point*

$$z = \frac{m_2 z_1 + m_1 z_2}{m_1 + m_2}, \quad m_1 m_2 \neq 0,$$

which divides the segment (z_1, z_2) in the real constant ratio $m_1 : m_2$. A necessary and sufficient condition that C be a curve is that C_1 and C_2 be segments of two parallel lines.

The sufficiency of the condition is immediate; all points z lie on a certain line parallel to C_1 and C_2. The locus C is connected since the loci C_1 and C_2 are connected (see I, p. 103). Hence C is a curve, a line segment.

To prove the necessity of the condition we resort to the type of reasoning just used in the proof of Theorem II. Let P be an arbitrary point of C_1, and C'_2 the curve obtained from C_2 by similarity transformation with center P. When P varies on C_1 and takes a position P', every point Q of C'_2 varies and takes a position Q' such that QQ' is parallel to PP'. According to our hypothesis on C, of which C'_2 is a part, and when we allow PP' to become smaller and smaller, we find that the tangent to C'_2 at Q is parallel to the tangent to C_1 at P. But Q is an arbitrary point of C'_2; then C'_2 (and consequently C_2) is a curve whose tangent has but a single direction and hence is a straight line segment. On the other

the plane of their circle. If under these conditions the locus of z_4 is not the whole of space, which is always the case if the arbitrary surfaces S_1, S_2, S_3 are sufficiently small and sufficiently remote from each other, simple consideration of the boundary of the locus of z_4 as in Theorem II gives us the following theorem of pure geometry which seems to be new:

The three surfaces S_1, S_2, S_3 define a congruence of circles C every circle of which cuts all those surfaces at points z_1, z_2, z_3 at the same angle, and such that if C is inverted into a straight line the planes tangent to S_1, S_2, S_3 at z_1, z_2, z_3 are all parallel. That is, any sphere through C cuts S_1, S_2, S_3 at z_1, z_2, z_3 at equal angles. If any surface S_4 is defined as the locus of points z_4 defined by a real constant cross ratio with z_1, z_2, z_3, the isogonal property holds for all the surfaces S_1, S_2, S_3, S_4.

The congruence C is a generalization of the well known normal congruence, for which the cross ratio property is also well known.

hand, C_1 is a curve whose tangent at an arbitrary point P is parallel to the line C_2. Hence C_1 and C_2 are segments of parallel lines.

Theorem III is not true for the degenerate values $m_1=0$ or $m_2=0$. The following theorem is likewise false for the degenerate values of the cross ratio:

THEOREM IV. *Let three curves C_1, C_2, C_3 be the loci of points z_1, z_2, z_3, respectively. Denote by C_4 the locus of the point z_4 defined by the real constant cross ratio*

$$\lambda=(z_1, z_2, z_3, z_4) \qquad \lambda \neq 0, 1, \text{ or } \infty .$$

A necessary and sufficient condition that C_4 be a curve is that C_1, C_2, C_3 be arcs of a single circle having no point common to all.

The sufficiency of the condition is immediate. If two of the points z_1, z_2, z_3 coincide, z_4 coincides with them and hence is on the circle of which C_1, C_2, C_3 are arcs. If no two of the original points coincide, z_4 is concyclic with those three points and hence also on that circle. The locus C_4 is connected (proved as in I, p. 103) and hence is an arc of that circle.

To demonstrate the necessity of the condition, we notice that C_1, C_2, C_3 must be circular arcs; otherwise we could choose suitably the point at infinity on C_1, for example, and C_2 and C_3 would become two curves not parallel lines, which would be in contradiction to Theorem III. If C_1, C_2, C_3 are not arcs of a single circle there is a point P which belongs to C_1, for example, but which does not belong to both of the circles of which C_2 and C_3 are arcs. When P is transformed to infinity, C_2 and C_3 are not both straight lines, so we have again a contradiction of Theorem III. If the arcs C_1, C_2, C_3 of a single circle have a point common to, all the locus of z_4 is the entire plane instead of a curve. The proof of Theorem IV is thus complete.

4. Theorem V, the interest of regions bounded by circular arcs. Theorems I-IV all indicate the central position in this study occupied by circles and circular arcs; this is heightened by the invariance of circles and circular arcs under linear transformation. The following theorem results from Theorem II:

THEOREM V. *Let the regions T_1, T_2, T_3, each bounded by a finite number of circular arcs, be the loci of points z_1, z_2, z_3, respectively. Then the region T_4 which is the locus of the point z_4 defined by the real constant cross ratio:*

$$\lambda=(z_1, z_2, z_3, z_4)$$

is also bounded by a finite number of circular arcs.

Every point of the boundary of T_4 corresponds to points z_1, z_2, z_3 on the boundaries of T_1, T_2, T_3 and such that the circle C through the four points cuts the boundaries of T_1, T_2, T_3 at vertices or such that the angles between C and the boundaries of T_1, T_2, T_3 are equal in magnitude. Corresponding to each set of circles, one a boundary (or in part forming part of the boundary) of each of the

original regions, there are four circles of the type of C_4 (Lemma IV, I, p. 105) according as C cuts all three circles at the same angle or a definite one at an angle supplementary to the angle cut on the other two.* Corresponding to each vertex of one of the original regions and two circles, one a boundary of each of the other of the original regions, there are two circles of the type of C_4 (Lemma IV, I, p. 105), and corresponding to two vertices of two separate original regions and a circle bounding the other region there is but one circle of this type. Since there are but a finite number of circular arcs and of vertices in the boundaries of each of the regions T_1, T_2, T_3, we have in all but a finite number of circles whose points can be boundary points of T_4; every boundary point of T_4 lies on at least one of these circles. The theorem follows at once. It is also true that if T_1, T_2, T_3 are bounded by at most a countable infinity of circular arcs, then T_4 is bounded by at most a countable infinity of circular arcs.

Regions of the sort considered in Theorem V are particularly interesting if they are bounded by *entire* circles instead of arcs of circles. We shall study in some detail the loci connected with such regions. The simplest case is that of circular regions, and has already been treated. The next simplest case is that of annular regions. We define an *annular region* to consist of the points common to two circular regions whose boundaries either have no point in common, or are tangent, or coincide.† An annular region is thus a closed region; every circular region is an annular region.

5. Theorem VI, the locus resulting from three circles; case of a single null circle. A general result concerning loci generated by regions whose boundaries are entire circles is to be proved later (Theorem VII). We now prove a preliminary result:

THEOREM VI. *If the loci of the points z_1, z_2, z_3 are the circles S_1, S_2, S_3, respectively, then the locus S_4 of the point z_4 defined by the real constant cross ratio*

$$\lambda = (z_1, z_2, z_3, z_4)$$

is an annular region.

For the values $\lambda = 0$, 1 or ∞, we must have at least two of the points z_1, z_2, z_3, z_4 coincident. Thus if $\lambda = 0$ the locus of z_4 is the entire plane or S_3 according as S_1 and S_2 have or have not a point in common. Similar facts hold for the other degenerate values of λ; compare I, p. 103. In the future we assume λ to have none of these values; then no two of the points z_1, z_2, z_3, z_4 coincide unless three of them coincide. Furthermore we place ourselves in what we may call the *general*

* This is true even if the three circles are coaxial; see the detailed proof of Lemma IV, I, p. 105 ff.

† It is possible to define annular regions so as to exclude the possibility of tangency of the bounding circles. But under this possible definition Theorem VI is no longer true; compare the results of §5.

situation, where none of the circles S_1, S_2, S_3 is a null circle, where no two of these circles are tangent, and where the three circles are not coaxial. We may immediately treat the case that the three original circles have a common point; the points z_1, z_2, z_3 may all be considered to lie at that point, z_4 may lie anywhere in the plane, so its locus is the entire plane.

The fact that S_4 is a two-dimensional continuum appears from the results of §3; the fact that the locus S_4 is connected appears by the reasoning used in I, p. 103.

Choose two arbitrary but distinct points P_1 and P_2 on S_1 and S_2, respectively, and let us consider the locus of z_4 defined by

$$(P_1, P_2, z_3, z_4) = \lambda$$

when the locus of z_3 is S_3. If we transform P_1 to infinity, z_4 is a point which divides the segment $P_2 z_3$ internally or externally in a constant ratio. Hence the locus of z_4 is a non-null circle S_4'. If P_1 and P_2 are chosen coincident, z_4 must coincide with them and its locus is a null circle. Let now P_2 trace the circle S_2, while P_1 remains fixed. If we determine the locus of all points of the variable circle S_4' we shall have precisely the locus of z_4 defined by

$$(P_1, z_2, z_3, z_4) = \lambda$$

when z_2 and z_3 have as their loci the circles S_2 and S_3.

We shall first assume that S_2 and S_3 are not coaxial with the null circle P_1 and that neither passes through P_1. The family of circles α through P_1 and cutting S_2 and S_3 at equal angles forms a coaxial family; compare the theorem quoted in I, p. 106. If P_1 is the point at infinity, this family passes through the external center of similitude of S_2 and S_3 if S_2 and S_3; are equal in size and P_1 is the point at infinity, this family is a family of parallel lines. Consider the circle α through the point P_2 and on α the point A determined by the cross ratio

$$(P_1, P_2, z_3, A) = \lambda,$$

where z_3 is the intersection of α with S_3 such that the tangent to S_3 at z_3 becomes parallel to the tangent to S_2 at P_2 when the circle α is transformed into a straight line. The point z_3 is uniquely determined by these conditions except in the particular case that α is orthogonal to S_2 and S_3; in this case we determine z_3 by continuity.

When the circle α varies, the point A traces a circle σ_1, by Lemma IV (I, p. 105). Moreover, the point A is continually on the circle S_4' corresponding to P_1 and P_2, and when α is transformed into a straight line, the line tangent to S_4' at A is parallel to the line tangent to S_3 at z_3. Hence as α varies, S_4' remains constantly tangent to σ_1; there is a circle S_4' tangent to σ_1 at every point of σ_1.

The family of circles β through P_1 and cutting S_2 and S_3 at supplementary angles also forms a coaxial family. When P_1 is the point at infinity the circles β all pass through the internal center of similitude of S_2 and S_3. When β varies, the procedure just employed for α can be used to show that the circle S_4' remains constantly tangent to a second circle σ_2 at a point of the circle β through P_2; there is a circle S_4' tangent to σ_2 at every point of σ_2.

The circles σ_1 and σ_2 are distinct. In fact, let us consider P_1 still as the point at infinity, and denote by A_2 and A_3 the centers of S_2 and S_3. Then the reasoning used in the proof of Lemma IV (I, p. 105) shows that σ_1 and σ_2 have as common center the point z_4 defined by

$$(P_1, A_2, A_3, z_4) = \lambda.$$

The external center of similitude of S_2 and S_3 (and which may be the point at infinity) is the external center of similitude or the internal center of similitude for *both* the pairs of circles σ_1, S_2 and σ_1, S_3; the internal center of similitude of S_2 and S_3 is the external center of similitude for one of the pairs of circles σ_2, S_2 and σ_2, S_3 and the internal center of similitude for the other pair. Neither of the circles σ_1 and σ_2 can coincide with either of the circles S_2 and S_3, for the common center of σ_1 and σ_2 is distinct from A_2 and A_3. It follows from this fact and the fact that the internal center of similitude of S_2 and S_3 differs from their external center of similitude that σ_1 and σ_2 are distinct.

The locus of the points of the variable circle S_4' is now apparent. Suppose for convenience in phraseology that σ_2 is interior to σ_1. The circle S_4' moves continuously so as to touch both σ_1 and σ_2 and touches every point of both circles. Either S_4' includes σ_2 or S_4' does not include σ_2, but in either case the locus of the points of S_4' is the annular region between and bounded by σ_1 and σ_2. In particular σ_2 (or σ_1) may be a null circle, but the locus of the points of S_4' is never the entire plane.

We have left aside the case that S_2 and S_3 have a common center M when P_1 is transformed to infinity. This situation will not be discussed in detail, but is entirely analogous to the case already treated. The two circles σ_1 and σ_2 have the common center M; if S_4 is chosen in any position and rotated about M, the two circles whose common center is M and which are tangent to S_4' are seen to play the rôle of σ_1 and σ_2. These two circles are the boundaries of the annular region which is the locus of the points of S_4'.

The case where S_2 or S_3, say for definiteness S_3, is a straight line remains to be considered. The circle S_4' is in every position a straight line parallel to S_3. When P_2 varies over S_2, the locus of the points of S_4' is seen to be a strip of the plane bounded by two lines parallel to S_3, which is an annular region. The theorem is then proved for the case that S_1 is a null circle. We notice that in

no case can σ_1 and σ_2 simultaneously be null circles, that they never coincide, and that they are tangent when and only when P_1 lies on S_2 or S_3.

6. Theorem VI, the locus resulting from three circles; general case. Our theorem, thus completely proved when S_1 is a null circle P_1, will be proved in the general case by determining the locus of the points of the annular region S_4'' bounded by σ_1 and σ_2 as the point P_1 traces the circle S_1. We shall suppose for the present that the three circles S_1, S_2, S_3 are not coaxial and that no two of them are tangent. It will follow that in Lemma IV (I, p. 105) we can say that as z_1 is made to vary continuously and in one sense on its locus, the points z_1, z_2, z_3, z_4 all vary continuously and in one sense on their loci. It should be remarked in connection with Lemma IV (I, p. 105) that it is essential to suppose the variable circle C and the points z_1, z_2, z_3 to vary continuously or that at least a suitable convention be made as to the choice of the points z_1, z_2, z_3 on C when C is orthogonal to C_1, C_2, C_3. For in that position of C, the condition that when C is transformed into a straight line the lines tangent to the three circles at z_1, z_2, z_3 are all parallel does not determine sufficiently these three points so that z_4 shall lie on C_4; and there are in general some extraneous points which enter into our locus if some convention concerning them is not made.

When P_1 varies and traces S_1, σ_1 varies also so as to remain constantly tangent to two fixed circles τ_1 and τ_2. For consider the circle C' through P_1 and which cuts S_1, S_2, S_3 all at the same angle. Choose on C' the points z_2 on S_2 and z_3 on S_3 such that when C' is transformed into a straight line the lines tangent to S_1, S_2, S_3 at P_1, z_2, z_3 are parallel. Then the point z_4 lies on σ_1 and also lies on the circle τ_1 of Lemma IV (I, p. 105) corresponding to the variable circle which cuts the three original circles at equal angles. Moreover when C' is transformed into a straight line the lines tangent to S_1 at P_1, S_2 at z_2, S_3 at z_3, σ_1 at z_4, τ_1 at z_4 are all parallel. Hence σ_1 is tangent to τ_1 at z_4.

In a precisely similar way it also appears that σ_1 is tangent to the circle τ_2 generated as described in Lemma IV (I, p. 105) corresponding to a variable circle C'' which cuts S_1 at an angle supplementary to the angles cut on S_2 and S_3. The variable circle σ_2 remains always tangent to two fixed circles ω_1 and ω_2 corresponding to a variable circle C''' which cuts S_2 at an angle supplementary to that cut on S_1 and S_3, and a variable circle C'''' which cuts S_3 at an angle supplementary to that cut on S_1 and S_2.

If the variable circle σ_1, is never a null circle, τ_1 and τ_2 do not intersect and are not tangent; for definiteness suppose τ_2 interior to τ_1. Whether σ_1 moves so as to include τ_2 or so as to exclude τ_2, σ_1 passes through every point of the plane between τ_1 and τ_2, so every such point is a point of S_4. A precisely similar remark obtains with reference to σ_2.

If σ_2 is ever a null circle, we shall prove that no arc of either of the circles

ω_1 and ω_2 can be a part of the boundary of S_4. The circles ω_1 and ω_2 must either intersect or be tangent. Suppose σ_2 never to pass through the point at infinity. When σ_2 is a null circle, σ_1 entirely surrounds σ_2. As σ_2 varies, immediately exterior to it there are always points of the final locus S_4, for σ_1 and σ_2 can never coincide and they vary continuously. It may occur that σ_1 and σ_2 become tangent, but this can happen only when P_1 is on S_2 or S_3 and thus only at the most at four isolated points. Except possibly at these four points, ω_1 and ω_2 are entirely embedded in points of S_4. A similar remark is evidently true for σ_1 and the circles τ_1 and τ_2, if σ_1 is ever a null circle.

The boundary of S_4 consists entirely of points of the circles τ_1, τ_2, ω_1, ω_2, by Theorem IV. The region S_4 is closed and its boundary cannot consist in whole or in part of isolated points. We may therefore suppose that at least one of the circles σ_1 or σ_2, say for definiteness σ_1, is never a null circle. If σ_1 passes through every point of the plane, S_4 is surely the whole plane. If σ_1 does not pass through every point of the plane, we suppose that it does not pass through the point at infinity. If σ_2 is ever exterior to σ_1, it always remains exterior; if ever interior, it always remains interior; for definiteness suppose that σ_2 always remains interior to σ_1. If σ_1 surrounds neither τ_1 nor τ_2, σ_1 passes through every point between τ_1 and τ_2, and every point of S_4'' is a point between those two circles, so S_4 is the annular region bounded by the non-intersecting circles τ_1 and τ_2.

If τ_2 is interior to τ_1 as well as to σ_1 (σ_2 constantly interior to σ_1), τ_1 is one boundary of S_4. But no arc of τ_2 can be a part of the boundary of S_4, since except at most for four isolated positions there are points of S_4'' interior and adjacent to σ_1 and hence there are points of S_4 interior and adjacent to τ_2. That is, not more than one of the circles τ_1 and τ_2 can be a part of the boundary of S_4; no arc of either circle is a part of that boundary unless that entire circle is a part of the boundary. The corresponding statement is true for ω_1 and ω_2.

In any case, then, the boundary of S_4 is composed of at most two of the circles τ_1, τ_2, ω_1, ω_2; it follows quite easily from our previous reasoning that these two circles do not intersect, so S_4 is an annular region.

The foregoing reasoning in proof of Theorem VI is not essentially altered and need not be further considered in detail if any of the circles τ_1, τ_2, ω_1, ω_2 is a null circle or if either of the circles σ_1 or σ_2 remains fixed during the motion of P_1. But on the other hand, it is necessary to note that we have essentially two distinct pairs of circles τ_1 and τ_2, and ω_1 and ω_2, or at least that when we speak of the circle σ_1 as tangent to τ_1 at z_4' and tangent to τ_2 at z_4'' we are not dealing with two points z_4' and z_4'' which always coincide and at the same time with two circles τ_1 and τ_2 which always coincide or are tangent at z_4' $(=z_4'')$. We prove this by choosing P_1 of such a nature that C' does not cut S_1 orthogonally and is not tangent to S_1, and such that σ_1 is not a null circle; such choice is possible under

our restrictions on S_1, S_2, S_3. Then C' and C'' cannot cut S_1 at the same angle and hence are distinct. The lines tangent to τ_1 (also to σ_1) at z_4' and to τ_2 (also to σ_1) at z_4'' cannot coincide, by the property of those tangent lines when C' and C'' are transformed into straight lines. Moreover σ_1 was chosen a non-null circle, so we are led to the conclusion that z_4' and z_4'' are distinct. This remark is similarly applicable to the points on the circles ω_1 and ω_2, so Theorem VI is completely proved under our assumption that S_1, S_2, S_3 are not coaxial and that no two of them are tangent.

For the more general case which makes no such assumption concerning S_1, S_2, S_3,* we consider variable auxiliary circles which approach the three given circles, and are such that at no stage are these three variable circles coaxial nor are any two of them tangent. The locus corresponding to the variable circles is always an annular region; it approaches uniformly the locus corresponding to the three given circles; this latter locus is therefore an annular region.

7. **Theorem VII, circular boundaries lead to circular boundaries; case of a single null circle.** We shall now make use of Theorem VI and its method of proof in the demonstration of our general theorem concerning loci whose boundaries are entire circles.

THEOREM VII. *If the loci of the points z_1, z_2, z_3 are, respectively, T_1, T_2, T_3, regions each bounded by a finite number of non-intersecting circles, the locus of z_4 defined by the real constant cross ratio*

$$\lambda = (z_1, z_2, z_3, z_4)$$

is a region T_4 also bounded by a finite number of non-intersecting circles. The number of circles bounding T_4 is not greater than the greatest number of circles bounding any of the regions T_1, T_2, T_3. In particular if T_1, T_2, T_3 are annular regions, T_4 is also an annular region; if T_1, T_2, T_3 are circular regions, T_4 is also a circular region.

In counting the number of boundaries of a region, any region T_i which consists merely of the points of a single circle is to be considered as having two boundaries. This is a natural convention, for we may think of T_i as the limit of a proper annular region as the two bounding circles approach each other. Some such convention is desirable so that the present theorem shall accord with Theorem VI.

It is of course true that T_4 may have as many bounding circles as the greatest number of circles bounding any of the regions T_1, T_2, T_3. This always occurs

* Theorem VI is extremely easy to prove for three coaxial circles. For three circles through two fixed points or all tangent at a single point the locus is the entire plane. For the other case the circles can be transformed so as to be made concentric. From symmetry about the common center and from the connectedness of the locus (as in I, p. 103) it follows that S_4 is an annular region.

($\lambda \neq 0, 1, \text{or } \infty$) if for example T_1 and T_2 are points; T_3 is the region or at least one of the regions with the greatest number of bounding circles. If T_1 and T_2 are sufficiently small, the number of their bounding circles has no effect; T_3 and T_4 have the same number of bounding circles.

The cases $\lambda = 0$, 1, or ∞ or that one of the given regions is the entire plane are to be treated as in I, p. 103, and will not be further considered. We shall further suppose for the present, that no two bounding circles of any one of the given regions are tangent, that no three circles, boundaries respectively of the three given regions, are coaxial, and that none of the three given regions is a point or a circle.

The boundary of the locus T_4 is composed of circles S_4 or arcs of these circles which are generated by the point z_4 determined by its cross ratio with the points z_1 on a circle S_1 which is a boundary of T_1, z_2 on a circle S_2 which is a boundary of T_2, and z_3 on a circle S_3 which is a boundary of T_3, while the circle S through these four points cuts S_1, S_2, S_3 all at the same angle or cuts one at an angle supplementary to the angle cut on the other two. We shall prove that *if an arc of any one of these circles S_4 is part of the boundary of T_4, the entire circle S_4 is part of the boundary of T_4.* For definiteness suppose S to cut S_1, S_2, S_3 all at the same angle so that the circle S_4 with which we are concerned is τ_1.

Since no two of the circles bounding T_1 intersect, there are points of T_1 adjacent to and all along one side of S_1, and similarly for the circles S_2 and S_3. We shall prove that even if we consider the entire circular region T_1' lying on that particular side of and bounded by S_1 as the locus of z_1 and the regions T_2' and T_3' similarly formed from S_2 and S_3 as the loci of z_2 and z_3, the locus of z_4 is that circular region T_4' bounded by τ_1 which contains τ_2, ω_1, ω_2. Since T_1' entirely contains T_1, T_2' entirely contains T_2, and T_3' entirely contains T_3, and since every point of τ_1 is a point of the locus of z_4 corresponding to the regions T_1, T_2, T_3, it follows that the entire circle τ_1 is a boundary of T_4; there are points z_4 corresponding to T_1, T_2, T_3 lying on one side of and all along τ_1, but there are no such points z_4 lying on the opposite side of τ_1.

The circle τ_1 is a part of the boundary of the annular region corresponding to the three circles S_1, S_2, S_3 as in Theorem VI; otherwise it is surely not part of the boundary for the regions T_1, T_2, T_3. Similarly, if we fix a point P_1 on S_1 we know that σ_1 must be a part of the boundary of the locus for P_1, S_2, S_3.

If we fix P_1 on S_1 and P_2 on S_2, we have points z_3 of T_3 on one side of and all along S_3, and hence we have points z_4 of T_4 on one side of and all along the circle S_4' of the proof of Theorem VI. Transform P_1 to infinity. When P_2 traces S_2, if S_4' always lies between σ_1 and σ_2 but surrounds neither, it passes through every point of the annular region between the circles. Moreover, if the points z_4 of T_4 just mentioned lie outside of S_4', no point of either σ_1 or σ_2 can be a point of the

boundary of T_4, and hence no arc of τ_1 can be a part of the boundary of T_4. Then the points z_4 which correspond to T_4' must all lie interior to S_4' and as P_2 varies on S_2 we have no new points added to the locus of z_4 because T_4 was replaced by T_4'.

When P_2 moves from S_2 into T_2, for example along the line L through P_2 cutting S_2 and S_3 at the same angle, S_4' moves so as to continue to cut L at the same angle, but is no longer tangent to σ_1 or σ_2. In fact, the intersections of S_4' with L are the points z_4 determined by the cross ratio of points P_1 (fixed), P_2 (variable on L), and the fixed intersections of L with S_3. Motion of P_2 causes these two intersections to vary. When P_2 is at either intersection of L with S_2, S_4' is tangent to σ_1 and σ_2; when P_2 moves on L from one of these intersections to the other in T_2', S_4' varies, but always varies so as to cut L at the same angle, which is the angle cut on L by σ_1. If we assume for definiteness that S_4' is interior to σ_1, we see that this motion of S_4' must always keep S_4' interior to σ_1, and σ_2 is not a part of the boundary of T_4 nor of T_4', and that the entire interior of σ_1 belongs to T_4'.

Precisely similar reasoning obtains if S_4' lies between σ_1 and σ_2 but encloses one circle and not the other, but we shall not give the details. We further omit the detailed treatment if S_2 or S_3 is a straight line (P_1 being at infinity). In every case we add no new boundary nor take away a part of the boundary of the locus of z_4 by replacing T_2 and T_3 by T_2' and T_3', so far as concerns the circle τ_1. Moreover the locus of z_4 in the latter case is not the whole plane unless the locus in the former case is the whole plane.

8. Theorem VII, circular boundaries lead to circular boundaries; general case. We shall now extend this reasoning by considering the locus of the points of the circle σ_1 as P_1 traces S_1. Suppose for definiteness that τ_1 is exterior to τ_2 and that σ_1 lies between the two circles but τ_2 is exterior to σ_1. We have assumed a part of τ_1 a part of the boundary of T_4, so the points of T_4 and hence of T_4' previously determined lie interior to σ_1. Similarly, a small motion of P_1 on S and into the interior of T_1 either moves σ_1 always interior to τ_1, or always exterior to τ_1; this follows from the continuity of the motion of σ_1 due to the continuous motion of P_1, and from the properties of the cross ratio determining the intersection of σ_1 with S which are shortly to be considered in detail. Since we are assuming at least an arc of τ_1 to be part of the boundary of T_4, it follows that such small motion of P_1 always moves σ_1 interior to τ_1.

Continuous motion of P_1 in one sense along S and in T_1' from one intersection of S with S_1 to the other intersection causes σ_1 to move. Each intersection of S with σ_1 is a point z_4 determined by its cross ratio with P_1, an intersection of S with S_2, and an intersection of S with S_3. The angle of intersection of S and σ_1 does not change, and this is the same as the angle of intersection of S and τ_1.

In its initial and final positions σ_1 is interior to τ_1; it therefore follows (in fact becomes evident if S is transformed into a straight line) that every point of σ_1 remains always interior to τ_1.

Determination of the locus of z_4 in this manner gives us the entire region T_4'; we have shown every point of T_4' to lie interior to or on τ_1. Every point interior to or on τ_1 is on a circle S and is interior to or on the circle σ_1 for some choice of P_1 in T_1' and hence T_4' is the interior (boundary included) of τ_1.

There is a possibility which could conceivably arise in this proof, namely that for some positions of S that circle does not cut S_1 and that for some point P_1 interior to T_1' but not on a circle S cutting S_1*, the circle σ_1 should be completely or partially exterior to τ_1; our reasoning as given does not permit this possibility. If there is one of the regions T_1', T_2', T_3' for which the circle S passes through every point, we change our notation if necessary so that that region shall be T_1', and our proof is valid as given. If there is no region T_1', T_2', T_3' through every point of which passes a circle S cutting S_1, S_2, S_3, it follows from the proof of Lemma IV (I, p. 105) that a point P_1 on S_1 but on neither S_2 nor S_3 can be chosen which lies in both T_2' and T_3'. It follows from the development of §7 that no point of either σ_1 or σ_2 can be a boundary point of T_4', so the possibility suggested need not be considered further.

The case where σ_1 lies between τ_1 and τ_2 but separates those two circles needs to be considered in detail, but only slight modifications in the reasoning are necessary and these are left to the reader. This completes the proof that τ_1 is the boundary of T_4' and hence part of the boundary of T_4. There are points z_4 of T_4 all along one side of and adjacent to τ_1, so no other bounding circle of T_4 (the entire circle necessarily part of the boundary of T_4) can intersect τ_1.

We have made the assumption that no two bounding circles of any one of the regions T_i are tangent, that no three circles, boundaries respectively of the three

* A similar possibility should have been pointed out in I, pp. 111–112 in proving that (notation of I) either the entire exterior or the entire interior of each of the circles C_4', C_4'', C_4''', C_4'''' belongs to the region C_4. For definiteness suppose this possibility to arise in connection with Lemma IV (I, p. 105), Case I; we consider the configuration simplified by transformation as in I, p. 106. The statement desired is evident if the regions C_1, C_2, C_3 are all interior to their bounding circles, so we may suppose one of these regions, say C_3, exterior to its bounding circle C_3. We wish to prove a point z_4 exterior to C_4' a point of the final envelope C_4 and need consider only points z_4 on circles C which do not cut C_1, C_2, C_3. Then there is a circle through z_4 which cuts C_1 and C_2 and which lies entirely in the region C_3. We may choose z_1 and z_2 on this circle and lastly z_3, so that these three points are in their proper envelopes and have the proper cross ratio with z_4.

The entire reasoning given in I, pp. 110–112, to prove that arcs of but one of the circles $C_4^{(j)}$ can be a part of the boundary of C_4 is no longer strictly necessary, for the proof of Theorem VII of the present paper contains a proof of Theorem I. Moreover, the reasoning of I, pp. 110–112, can be replaced by the more simple and elegant argument used in §6 (or even §15) of a paper by the writer shortly to appear in these Transactions.

given regions, are coaxial, and that none of the given regions is a point or a circle. To extend our result to include these special cases, we consider a sequence of sets of three regions which satisfy our restrictions but which approach the more special given regions; such an auxiliary sequence can always be constructed. Then the limit of the locus of z_4 for the variable regions is the locus of z_4 for the limit regions, from which it follows that for the limit case the locus T_4 is bounded by a finite number of entire circles.

9. Theorem VII, number of circular boundaries. It remains for us to prove that the number of circles bounding T_4 is not greater than the greatest number of circles bounding any of the regions T_1, T_2, T_3. In the detailed proof we restrict the boundaries of the original regions as before; this restriction is raised by the limiting process previously used and need not be mentioned further.

We keep P_1 (z_1 in any position) fixed for the moment, and suppose that the point at infinity is not a point of the locus T_4'' corresponding to the points z_2 of T_2 and z_3 of T_3. Consider the mechanism used in the proof of Theorem VI; first suppose S_4' to lie between σ_1 and σ_2 and σ_2 to be interior to σ_1 but not interior to S_4'. Under the assumption that σ_1 is part of the boundary of T_4'', when P_2 traces S_2, S_4' moves tangent to σ_1 and σ_2. If Σ_4' is a circle which is obtained from the proper cross ratio from P_1, P_2, and the points of a bounding circle Σ_3 of T_3 other than S_3, then Σ_4' is interior to S_4'. During the motion of P_2 on S_2, Σ_4' cannot trace a boundary of T_4''. Moreover, if P_2 is moved continuously, Σ_4' moves continuously; Σ_4' can never be made to surround a lacuna in T_4'', from which it follows that Σ_3 in combination with any bounding circle of T_2 cannot lead to a boundary of T_4''. Bounding circles of T_4'' can proceed only from combination of S_3 with bounding circles of T_2, so the number of bounding circles of T_4'' is not greater than the number of bounding circles of T_2.

Second, suppose S_4' to lie between σ_1 and σ_2 and σ_2 to be interior to both σ_1 and S_4'. When σ_1 is supposed as before part of the boundary of T_4'', the circle Σ_4' must as before be interior to S_4'. There is a circle Σ_4' interior to S_4' corresponding to each of the m bounding circles Σ_3 of T_3 (other than S_3), and in the region traced out by these circles as P_2 moves around S_2 there are at most m lacunae, each constantly surrounded by a circle Σ_4'. When P_2 is made to trace the whole of T_2, one or more of these lacunae may disappear, but none of them can divide to make two or more lacunae in T_4''. It follows that the number of bounding circles of T_4'' is not greater than the number of bounding circles of T_3.

The case that σ_2 is a point leads simply to one bounding circle of T_4'', so in every case T_4'' is bounded by a number of circles not greater than the greater number of bounding circles of T_2 and T_3.

If the reasoning just used is again applied regarding the locus of T_4'' as P_1 varies over T_1, it is seen that T_4 is bounded by a number of circles not greater

than the greater number of bounding circles of T_1 and T_4'', which is not greater than the greatest number of bounding circles of T_1, T_2, T_3, so the proof of Theorem VII is complete. It is also true that if T_1, T_2, T_3 are each bounded by at most a countable infinity of non-intersecting circles then T_4 is also bounded by at most a countable infinity of non-intersecting circles.

10. Theorem VII, ruler-and-compass construction for boundaries. As in Theorem I, the circles which bound the region T_4 of Theorem VII can be constructed by ruler and compass whenever λ is rational or is given geometrically; indeed this follows from the fact that any circle C_4 of Lemma IV (I, p. 105) can be so constructed; see I, p. 105, footnote. We also have a test for determining whether or not a given circle S_4 (in the notation of §7) is actually a part of the boundary of T_4. If none of the regions T_1, T_2, T_3 reduces to the points of a circle, we need merely suppose the circles S_1, S_2, S_3 to bound entire circular regions which lie on the same side of those circles as the given regions; then S_4 is a boundary of T_4 when and only when the locus of z_4 in the simplified situation is a circular region bounded by S_4. A test for this latter fact has already been determined in I, p. 112. If any of the regions T_1, T_2, T_3 reduces to the points of a circle, we may determine in what sense to consider the corresponding circular regions to lie by an investigation such as that of §7.*

To be sure, the test for determining whether S_4 can be a boundary of T_4 was proved only under certain restrictions on the bounding circles of T_1, T_2, T_3, but the test can be applied whether those restrictions are satisfied or not. Thus, if the circular regions which lie on the same side of S_1, S_2, S_3 as do the given regions T_1, T_2, T_3 lead to a region T_4' bounded by a circle τ_1 and which is not the whole plane, the original regions lead to a region T_4 bounded by the entire circle τ_1 and which is not the whole plane. On the other hand, if a circle S_4 is actually a boundary of T_4, there must be a circle which bounds the locus of z_4 of the approximating sequence and which approaches S_4. There are therefore entire variable circular regions T_1', T_2', T_3' corresponding which lead to a locus of z_4 not the whole plane; these approach entire circular regions formed in the manner described from certain boundaries of T_1, T_2, T_3, and which lead to a locus of z_4 not the whole plane. It follows that the test to determine whether S_4 is a boundary of T_4 is valid in every case. If any of the given regions is a circle, the results in the given case may also be indicated by approximating non-special cases.

* One particular case of three circles S_1, S_2, S_3 can be treated directly, namely, where one of those circles separates the intersections of the other two; compare I, p. 111. The locus of z_4 is always the entire plane. For any given point P of the plane there can be determined points z_1, z_2, z_3 on S_1, S_2, S_3 such that their cross ratio with P has a value very small and positive; these points can be moved continuously on their respective circles so that this cross ratio increases in value and becomes less than but nearly equal to unity. Then these points can be chosen so that λ takes any value between zero and unity; similarly, so that it takes any real value.

Great care should be used in determining loci and limits of loci of this nature if boundaries of the three given regions are tangent at a single point, for it may be merely by virtue of the fact that when z_1, z_2, z_3 coincide z_4 is undetermined that the locus of z_4 is the entire plane. Thus, if two finite points z_1 and z_2 have as their common locus a half-plane, the locus of the mid point z_4 of their segment is also that half-plane. When we consider not the problem of the ordinary ratio but that of the *cross ratio* of the two given points and the mid point of their segment with the point at infinity we are compelled to admit the point at infinity as a possibility for z_1 and z_2 and that the entire plane is the locus of z_4 (compare I, p. 111). For there can be found two variable points z_1 and z_2 such that z_4 is the mid point of their segment and which approach the point at infinity. When these two points coincide at infinity we ought therefore to consider that the four limit points have the proper cross ratio. But for an arbitrary point z_4 of the plane there cannot be found two variable finite points z_1 and z_2 *in their proper loci*, such that z_4 is the mid point of their segment and which approach the point at infinity. This phenomenon can occur only if the three regions have but a single common point and if the respective boundaries are tangent at that point.

This difference of behavior of ratios and cross ratios corresponds indeed to a difference of behavior in our original problem of the roots of the jacobian, as compared with the problem of the roots of the derivative of a polynomial. For most purposes these problems are equivalent if one of the ground forms has all its roots at infinity and the other has as its roots the roots of the polynomial considered. If all the roots of both ground forms coincide, the jacobian vanishes identically. Hence if in Theorem III (I, p. 112) the loci of m roots of f_1 and the remaining p_1-m roots of f_1 are two coincident half-planes, and if f_2 has all its roots coincident at infinity, the locus of the roots of the jacobian is the entire plane. But if these two coincident half-planes are the loci of m roots of a polynomial and the remaining p_1-m roots of that polynomial, and if the polynomial is not allowed to have infinite roots, the locus of the roots of the derivative of this polynomial is this same half-plane instead of the entire plane.

We have assumed in Theorem VII not only that the boundaries of T_1, T_2, T_3 are entire circles but also that no two of the bounding circles of one of these regions cut each other. If this latter part of the hypothesis is omitted, it is not true that T_4 is bounded by entire circles. Consider for example T_1 the point at infinity, T_2 the two finite crescents bounded by two intersecting circles S_2' and S_2'', and T_3 the interior and boundary of a circle S_3 whose center is A and whose radius is small in comparison with the radii of S_2' and S_2''. Choose λ so that when z_1 is at infinity z_4 lies midway between the finite points z_2 and z_3.

The locus of z_4 may be determined by fixing z_2, determining the locus of z_4 while z_3 varies over T_3, and then allowing z_2 to vary over T_2. That is, we shrink T_2 toward A as center of similitude in the ratio $1:2$, and determine the locus of the

points of a small circle whose radius is half the radius of S_3 and whose center varies all over this new configuration. The locus T_4 is therefore bounded by arcs of four different circles; no entire circle is a part of its boundary.

11. Successive application of Theorem I in determination of loci. An evident way of generalizing Theorem I is by successive application of that theorem. Thus, suppose we have n sets of the circular regions of Theorem I, $C_k^{(i)}$ $(k=1, 2, 3, 4; i=1, 2, \ldots, n)$. Any point z_4 which corresponds to points $z_k (k=1, 2, 3)$ in all the regions $C_k^{(i)}$ $(i=1, 2, \ldots, n)$ is located in all the regions $C_4^{(i)}$ $(i=1, 2, \ldots, n)$, so the region T_4 which is the locus of points z_4 determined by points z_k in the regions T_k common to all the $C_k^{(i)}$ $(i=1, 2, \ldots, n; k=1, 2, 3)$ is contained in all the regions $C_4^{(i)}$. But ordinarily T_4 will not be the entire region common to this last set of regions. A simple example is the case, $\lambda=\frac{1}{2}$:

$$\begin{array}{lll} C_1^{(1)} : \infty & C_1^{(2)} : \infty & T_1 : \infty \\ C_2^{(1)} : |z_2+2| \leqq 1 & C_2^{(2)} : -2 & T_2 : -2 \\ C_3^{(1)} : 2 & C_3^{(2)} : |z_3-2| \leqq 1 & T_3 : 2 \\ C_4^{(1)} : |z_4| \leqq \frac{1}{2} & C_4^{(2)} : |z_4| \leqq \frac{1}{2} & T_4 : 0. \end{array}$$

A particularly interesting case such that T_4 is the entire region common to the $C_4^{(i)}$ occurs when all the sets of circles $C_1^{(i)}, C_2^{(i)}, C_3^{(i)}, C_4^{(i)}$ come under a single case (e. g., Case I) of Lemma IV (I, p. 105), where these circles have a single coaxial system of circles C cutting them all at the same angle or a definite set or sets at angles supplementary to the angles cut on the other sets, and where all the circular regions $C_k^{(i)}$ for one value of i have the same disposition with respect to their bounding circles as the corresponding regions for every other value of i. It follows from the reasoning of I, pp. 111–112, as supplemented in a footnote to §8 of the present paper, that every point of the region common to all the $C_4^{(i)}$ is a point of the locus of z_4, so that common region is precisely T_4. The circle C then cuts the boundaries of T_1, T_2, T_3, T_4 all at the same angle or one or two of those regions at an angle supplementary to the angle cut on the others. Our notation here has supposed the sets of circles $C_k^{(i)}$ finite in number, but of course that is unnecessary.

A special case of this last result will be considered in some detail, where a set of regions $C_k^{(i)}$ is one and the same point for all values of i. We first state explicitly a special case of Theorem I, which corresponds to a theorem given in I, p. 115, and which is proved explicitly in S.

THEOREM VIII. *If the loci of the points z_1 and z_2 are the interiors and boundaries of the circles C_1 and C_2 whose centers are a_1 and a_2 and radii r_1 and r_2, respectively, then the locus of the point*

$$z = \frac{m_2 z_1 + m_1 z_2}{m_1 + m_2}$$

which divides the segment (z_1, z_2) in the real constant ratio $m_1 : m_2$ $(m_1 m_2 > 0)$ is the interior and boundary of the circle C whose center is

$$\frac{m_2 a_1 + m_1 a_2}{m_1 + m_2}$$

and radius

$$\frac{m_2 r_1 + m_1 r_2}{m_1 + m_2}.$$

The point

$$\frac{r_2 a_1 - r_1 a_2}{r_2 - r_1}$$

(which may be the point at infinity) is an external center of similitude for any pair of the circles, C_1 C_2, C.

We next prove a generalization of Theorem VIII:

THEOREM IX. *If two finite or infinite convex regions C_1 and C_2 having an external center of similitude P (which may be the point at infinity) are the loci of two points z_1 and z_2, respectively, then the locus of the point z which divides the segment (z_1, z_2) in the real constant ratio $m_1 : m_2$ $(m_1 m_2 > 0)$ is a region C such that P is an external center of similitude for any pair of the regions C_1, C_2, C.*

In Theorem IX, we tacitly assume C_1 and C_2 to be closed regions bounded by regular curves. We formulate the proof of the theorem for the case that P is finite, but a very slight change in phraseology will give the proof when P is infinite.

Consider a line L through P cutting the boundary of C_1 in points A_1 and A_1' and the boundary of C_2 in corresponding points A_2 and A_2'. If C_1 and C_2 are infinite, A_1' and A_2' are to be considered the point at infinity. When L rotates about P, the points A and A' which divide the segments (A_1, A_2) and (A_1', A_2'), respectively, in the ratio $m_1 : m_2$ trace the boundary of a convex region C such that P is an external center of similitude for any pair of the regions C_1, C_2, C. We shall prove C to be the locus of z. Any point z of C is a point of the locus, for we need merely choose z_1 and z_2 on the line Pz and having the same relative situation in C_1 and C_2 as z in C. In order to prove that every point z is in C we shall use a preliminary theorem which is a special case of Theorem IX and a limiting case of Theorem VIII:

THEOREM X. *If two half-planes π_1 and π_2 bounded by two parallel lines λ_1 and λ_2 and lying on the same side of those lines are the loci of two points z_1 and z_2, respectively, then the locus of the point z which divides the segment (z_1, z_2) in the real constant ratio $m_1 : m_2$ $(m_1 m_2 > 0)$ is a half-plane π bounded by a line λ parallel to λ_1 and λ_2; π lies on the same side of λ as π_1 and π_2 of λ_1 and λ_2. Any line which*

cuts $\lambda_1,\lambda_2,\lambda$ *cuts them in points* z_1, z_2, z *such that* z *divides the segment* (z_1, z_2) *in the ratio* $m_1:m_2$.

We prove Theorem X in the same spirit as Theorem VIII was proved in S. We use rectangular coördinates $z_1=x_1+iy_1$, $z_2=x_2+iy_2$, $z=x+iy$, and suppose the regions π_1 and π_2 to be $x_1 \geqq a_1$ and $x_2 \geqq a_2$, respectively. The region π is

$$x \geqq \frac{m_2a_1+m_1a_2}{m_1+m_2}.$$

For given z_1 and z_2 in their proper loci, the formula

$$x=\frac{m_2x_1+m_1x_2}{m_1+m_2}$$

shows at once that z is in π. If z is given in π we have only to choose

$$y_1=y_2=y,\quad x_1=a_1-\frac{m_2a_1+m_1a_2}{m_1+m_2}+x,\ x_2=a_2-\frac{m_2a_1+m_1a_2}{m_1+m_2}+x,$$

and we shall have z_1 and z_2 in π_1 and π_2 and such that z divides their segment in the ratio $m_1:m_2$. The axis of reals cuts λ_1, λ_2, λ in points z_1, z_2, z such that z divides the segment (z_1, z_2) in the ratio $m_1:m_2$, and hence that is true of any other line which cuts λ_1, λ_2, λ.

Theorem X is to be used in proving that every point z of Theorem IX lies in C. Through the point A_1 draw a line λ_1 which does not cut C_1 and through A_2 a parallel line λ_2, where A_1 and A_2 have the same significance as before. There is a half-plane π_1 bounded by λ_1 which contains all of C_1 and a corresponding half-plane π_2 bounded by λ_2 which contains all of C_2. The corresponding half-plane π bounded by the line λ through A contains all of C, from which it follows that any point z of Theorem IX lies in π. This is always true as L rotates about P and as λ_1, λ_2, λ envelop C_1, C_2, C_3, so z lies in C and Theorem IX is completely proved.

Our proof depends essentially on the fact that C_1 and hence also C_2 and C are convex, since we have passed a line λ_1 through an arbitrary point A_1 of the boundary of C_1 and supposed C_1 to lie entirely on one side of that line. If C_1 is not convex not only does the proof break down, but the theorem is false and we can even show that *the locus of* z *cannot be the region* C found as previously described. There are two points U_1 and V_1 on the boundary of C_1 such that no point of the segment U_1V_1 belongs to C_1. The analogous segment U_2V_2 has no point in C_2. Denote by U and V the points on the boundary of C which divide the segments U_1U_2 and V_1V_2 in the ratio $m_1:m_2$; the segment UV has no point in C. The lines U_1V_1, U_2V_2, UV are all parallel and the point z which divides the segment U_2V_1 in the ratio $m_1:m_2$ lies on the segment UV and hence is exterior to the region C.

We terminate our study of the loci connected with cross ratios by stating without further proof the theorem analogous to Theorem IX if $m_1m_2 < 0$:

THEOREM XI. *If two finite or infinite convex regions C_1 and C_2 having a finite internal center of similitude P are the loci of two points z_1 and z_2, respectively, then the locus of the point z which divides the segment (z_1, z_2) in the real constant ratio $m_1 : m_2$ $(m_1m_2 < 0)$ is a region C such that P is an external center of similitude for C_1 and C (or C_2 and C) and an internal center of similitude for C_2 and C (or C_1 and C).*

If C_1 and C_2 are not convex but satisfy the other conditions of the theorem, the corresponding region C cannot be the locus of z.

12. Application to the location of the roots of the jacobian of two binary forms. Theorem I was originally proved in I not for its own sake but for application to the study of the location of the roots of the jacobian of two binary forms. Thus there can be proved (I, p. 114):

THEOREM XII. *Let f_1 and f_2 be binary forms of degrees p_1 and p_2, respectively, and let the circular regions C_1, C_2, C_3 be the respective loci of m roots of f_1, the remaining p_1-m roots of f_1, and all the roots of f_2. Denote by C_4 the circular region which is the locus of points z_4 such that*

$$(z_1, z_2, z_3, z_4) = \frac{p_1}{m}$$

when z_1, z_2, z_3 have the respective loci C_1, C_2, C_3. Then the locus of the roots of the jacobian of f_1 and f_2 is composed of the region C_4 together with the regions C_1, C_2, C_3, except that among the latter the corresponding region is to be omitted if any of the numbers m, p_1-m, p_2 is unity. If a region C_i $(i=1, 2, 3, 4)$ has no point in common with any other of those regions which is a part of the locus of the roots of the jacobian, it contains precisely $m-1$, p_1-m-1, p_2-1, or 1 of those roots according as $i=1, 2, 3,$ or 4.

Theorem XII can be generalized so as to give applications for some of the results of the present paper. Thus we have directly:

THEOREM XIII. *Let f_1 and f_2 be binary forms of degrees p_1 and p_2, respectively, and let regions T_1, T_2, T_3 be the loci of m roots of f_1, the remaining p_1-m roots of f_1, and all the roots of f_2, these three sets of roots always to be contained in three variable circular regions all points of which are points of T_1, T_2, T_3, respectively. Denote by T_4 the locus of points z_4 such that*

$$(z_1, z_2, z_3, z_4) = \frac{p_1}{m}$$

when z_1, z_2, z_3 have the respective loci T_1, T_2, T_3. Then the locus of the roots of the jacobian of f_1 and f_2 is composed of the region T_4 together with the regions T_1, T_2, T_3, except that among the latter the corresponding region is to be omitted if any of the

numbers m, p_1-m, p_2 *is unity. If a region* T_i $(i=1, 2, 3, 4)$ *has no point in common with any other of those regions which is a part of the locus of the roots of the jacobian, it contains precisely* $m-1$, p_1-m-1, p_2-1, *or* 1 *of those roots according as* $i=1$, 2, 3, *or* 4.

It is in connection with Theorem XIII that most of the results of the present paper are to be considered so far as concerns their application to the location to the roots of the jacobian. It should be noticed, however, that Theorem XIII gives no better indication of the location of the roots of the jacobian of particular fixed forms than does Theorem XII, but may give a better indication than Theorem XII of the location of the roots of the jacobian of two forms whose roots vary in particular ways. The results of §11 can also be used independently of Theorem XIII in establishing slight extensions of Theorem XII which do apply to the jacobian of two fixed forms; this application will be made in a later paper.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

ON TRANSFORMATIONS WITH INVARIANT POINTS*

BY

J. W. ALEXANDER

1. There is a well known theorem by Brouwer† to the effect that every sense-preserving transformation of a spherical surface into itself such that both the direct transformation and its inverse are one-valued leaves invariant at least one point of the surface. It is proposed to prove a somewhat more general theorem by a simplified process and to outline one or two applications of the theorem in its extended form. The theorem and its proof both generalize automatically to any number of dimensions.

The results contained in this paper were obtained by the author a number of years ago but never published. Later, a more concise method of presentation suggested itself and made it possible to condense the argument into the form given below. Analogous theorems have recently been worked out independently by Professors Birkhoff and Kellogg as a first step towards the solution of some important problems on functional transformations.‡

2. **The transformation $S_1 : S_2$.** Let S_1 be the sphere

$$x^2 + y^2 + z^2 = 1$$

in real 3-space. It will be convenient to express the coördinates of a point P_1 of S_1 by means of two parameters u and v, such, for example, as the angles of latitude and longitude in ordinary polar coördinates. By an image S_2 of the sphere S_1 will be meant the locus in 3-space of a point $P_2(u, v)$ which varies continuously with the point $P_1(u, v)$ of S_1. The image S_2 will thus be defined by certain parametric equations

$$x = X(u, v), \quad y = Y(u, v), \quad z = Z(u, v). \tag{1}$$

We shall not exclude the case where the transformation determined by (1) carries two or more points of S_1 into the same point (x, y, z); in fact, in an ex-

* Presented to the Society, September 8, 1921.

† L. E. J. Brouwer. *Ueber eineindeutige, stetige Transformationen von Flächen in sich*, Mathematische Annalen, vol. 69 (1910). Also, Proceedings of the Section of Sciences Royal Academy at Amsterdam, vol. 13 (1911).

‡ G. D. Birkhoff and O. D. Kellogg, *Invariant points in function space*, these Transactions, vol. 23 (1922).

treme case, it may carry the entire surface S_1 into a single point. However, we shall make a convention similar to the one adopted in the theory of Riemann surfaces to the effect that a point of S_2 is not determined by position alone but also by association with a definite point of S_1, as determined by the relations (1). With this extension of language, there will be exactly one point of S_2 corresponding to each point of S_1.

3. **The index of the transformation $S_1 : S_2$.** Let us assume provisionally that the image S_2 of S_1 consists of a finite number of analytic pieces, so that the Gaussian integral

$$(2) \qquad \iint \begin{vmatrix} x & y & z \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{vmatrix} \frac{dudv}{r^3}, \qquad r = +\sqrt{x^2+y^2+z^2},$$

evaluated over S_2 has a meaning.* It is not difficult to see that this integral measures the solid angle subtended at the origin by the surface S_2 when the customary conventions are adopted as to the sign of the angle subtended by each surface element. For our present purposes, we merely notice that the conditions of integrability

$$(3) \qquad \frac{\partial}{\partial x}\left(\frac{x}{r^3}\right) + \frac{\partial}{\partial y}\left(\frac{y}{r^3}\right) + \frac{\partial}{\partial z}\left(\frac{z}{r^3}\right) = 0$$

are satisfied identically except at the origin and, consequently, that the value of the integral (2) remains unaltered when the surface S_2 is deformed continuously without meeting the origin, provided, of course, that the surface remains sufficiently regular so that the integral does not become meaningless. If S_2 crosses the origin, however, the integrand undergoes a discontinuity, and the integral changes abruptly by $\pm 4\pi$, the value obtained by integrating (2) over a small sphere with center at the origin. Evidently, therefore, the value of the integral over S_2 is $4k\pi$, where k is some integer denoting the difference between the number of times that a general ray issuing from the origin cuts the surface from the negative to the positive side and the number of times that it cuts the surface from the positive to the negative side.

In particular, the value of (2) integrated over S_1 itself is $\pm 4\pi$. We shall select the parameters u and v in such a way as to make this last number positive. The quantity k determined by the integration of (2) over S_2 for the same choice

* Cf. the note by Hadamard appended to Tannery's *Introduction à la Théorie des Fonctions d'une Variable.*

of parameters is then a function of the transformation $S_1 : S_2$ which remains invariant under a continuous deformation of S_2 which does not carry S_2 into contact with the origin. We shall call k the *index* of the transformation.

Since the conditions of integrability (3) are satisfied, we may do away with the restriction that the surface S_2 be composed of a finite number of analytic pieces and give the integral (2) a meaning, by extension, over any image S_2 of S_1 which does not meet the origin. For we shall prove that S_2 may always be approximated to any degree of accuracy by a second image S_2' of the analytic type and that the values of (2) for any two sufficiently good approximations of S_2 will always be the same. We may therefore say, by definition, that the value of the integral (2) over a general image S_2 is the same as its value over a sufficiently good approximating surface S_2'.

One obvious way of setting up the approximating surfaces S_2' is the following. Let α be any positive number less than the distance from the origin to the nearest point of S_2 to the origin. Then, owing to the uniform continuity of the transformation $S_1 : S_2$, it will be possible to cut up the sphere S_1 into a finite system of non-overlapping spherical triangles such that the images in S_2 of any two points from the interior and boundary of the same triangle are always within a distance of $\alpha/8$ from one another. Suppose, in the first instance, that the images of no three vertices of a spherical triangle are collinear. We may then approximate the image S_2 of S_1 by a second image S_2' consisting of plane triangular pieces, one corresponding to each spherical triangle of S_1 and such that the vertices of each plane triangle coincide with the images of the vertices of the corresponding spherical traingle. We thereby obtain an approximating surface S_2' composed of triangular regions (which may, of course, cut into one another) and such that each point of S_2' is within a distance of $\alpha/4$ from the corresponding point of S_2. When the images of the three vertices of a spherical triangle of S_1 are collinear, we may replace each of the images by a nearby point within a distance $\alpha/16$ of the original and so locate the new points that no three are collinear. We then proceed as before and build up an approximating surface S_2' out of triangles with vertices at the new points. This time, each point of S_2' will be within a distance of $\alpha/2$ from the corresponding point of S_2. It is therefore possible in every case to approximate the image S_2 to any degree of accuracy by a second image S_2' made up of analytic pieces (triangular regions, in fact).

Finally, the value of the integral (2) over two surfaces T and U which approximate the surface S_2 to within the accuracy of the surface S_2' defined above must be the same. For each point (x', y', z') of T is within a distance α of the corresponding point (x'', y'', z'') of U; consequently, the surface T may be de-

formed into the surface U without crossing the origin. Such a deformation could be defined, for instance, by the equations

$$x = \lambda x'' + (1 - \lambda)x', \quad y = \lambda y'' + (1 - \lambda)y', \quad z = \lambda z'' + (1 - \lambda)z',$$

where λ is allowed to vary from 0 to 1.

It follows at once from the above that we may define the index of any transformation $S_1 : S_2$ provided S_2 does not meet the origin, and that the index remains invariant under the most general continuous deformation of S_2 which does not bring this surface into contact with the origin.

4. The extended form of Brouwer's theorem. We shall be concerned more particularly with the case where every point of the image S_2 coincides with a point of S_1, so that the transformation $S_1 : S_2$ carries the sphere S_1 into itself or a part of itself. Such a transformation is extremely general: it may have any index, and its inverse need neither be one-to-one or even everywhere defined on the sphere. It will be shown that

The transformation $S_1 : S_2$ must have an invariant point if its index is not equal to -1.

For let the point (x_1, y_1, z_1) of S_1 be paired with the point (x_2, y_2, z_2) of S_2 under the transformation $S_1 : S_2$. Then, if we put

$$x = x_1 - x_2, \quad y = y_1 - y_2, \quad z = z_1 - z_2, \quad r = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2},$$

the integral (2) assumes the form

$$\iint \begin{vmatrix} x_1 - x_2 & y_1 - y_2 & z_1 - z_2 \\ \dfrac{\partial}{\partial u}(x_1 - x_2) & \dfrac{\partial}{\partial u}(y_1 - y_2) & \dfrac{\partial}{\partial u}(z_1 - z_2) \\ \dfrac{\partial}{\partial v}(x_1 - x_2) & \dfrac{\partial}{\partial v}(y_1 - y_2) & \dfrac{\partial}{\partial v}(z_1 - z_2) \end{vmatrix} \frac{du\,dv}{r^3}, \tag{4}$$

which has a perfectly definite meaning provided r does not vanish, that is, provided the transformation $S_1 : S_2$ has no fixed point. Furthermore, (4) remains invariant under any continuous deformation of either S_1 or S_2 which does not bring a point of one surface into coincidence with its image on the other.

Now, consider the following two deformations:

(*i*) S_1 is held fixed but S_2 is shrunk into coincidence with the origin by a radial contraction. Then, in the limit, the integral (4) reduces to

$$\iint \begin{vmatrix} x_1 & y_1 & z_1 \\ \dfrac{\partial x_1}{\partial u} & \dfrac{\partial y_1}{\partial u} & \dfrac{\partial z_1}{\partial u} \\ \dfrac{\partial x_1}{\partial v} & \dfrac{\partial y_1}{\partial v} & \dfrac{\partial z_1}{\partial v} \end{vmatrix} \frac{du\,dv}{r^3} = 4\pi. \tag{5}$$

(*ii*) S_2 is held fixed but S_1 is shrunk into coincidence with the origin. The integral (4) then becomes

$$(6) \qquad -\int\int \begin{vmatrix} x_2 & y_2 & z_2 \\ \frac{\partial x_2}{\partial u} & \frac{\partial y_2}{\partial u} & \frac{\partial z_2}{\partial u} \\ \frac{\partial x_2}{\partial v} & \frac{\partial y_2}{\partial v} & \frac{\partial z_2}{\partial v} \end{vmatrix} \frac{du\,dv}{r^3} = -4k\pi,$$

where k is the index of the transformation $S_1 : S_2$. Equating (5) and (6), we have at once $k = -1$ which proves that if there is no fixed point the index of the transformation must be -1.

5. Generalization to n dimensions. Evidently, both theorem and proof generalize at once to any number of dimensions by the use of the integral

$$(7) \qquad \int \cdots \int_{(n)} \begin{vmatrix} x_0 & \cdots\cdots & x_n \\ \frac{\partial x_0}{\partial u_1} & \cdots\cdots & \frac{\partial x_n}{\partial u_1} \\ \cdots & \cdots\cdots & \cdots \\ \frac{\partial x_0}{\partial u_n} & \cdots\cdots & \frac{\partial x_n}{\partial u_n} \end{vmatrix} \frac{du_1 \cdots du_n}{r^{n+1}} = 2^n k\pi.$$

It will be noticed, however, that the negative sign before the integral (6) will reappear in the generalizations to even values of n only, since it depends on the fact that the determinant in the integrand has an odd number of rows and columns. The form of the generalized theorem is, therefore, as follows:

Any one-valued continuous transformation $S_1 : S_2$ of an n-sphere

$$x_0^2 + x_1^2 + \cdots + x_n^2 = 1$$

into itself has an invariant point provided the index of the transformation is not $(-1)^{n+1}$.

6. Applications. Certain other theorems by Brouwer may now be proved in somewhat extended form.

Every one-valued continuous transformation of real projective n-space of even dimensions into itself or a part of itself has at least one invariant point. (In this and in the following theorem, the inverse transformation need not be single-valued.)

This theorem follows at once from the homeomorphism existing between the points of projective n-space and the point pairs of an n-sphere forming the extremities of the diameters of the n-sphere. (To set up the homeomorphism, we need only project the n-sphere from its center upon a projective n-plane.)

To every transformation of projective n-space into itself, there evidently correspond two transformations of the n-sphere into itself differing from one another by a reflection in the center of the sphere. But, if the number n is even, such a reflection changes the sign of the index, so that one of the two transformations surely has an index different from $(-1)^{n+1}$. This transformation therefore has a fixed point belonging to a fixed point pair, which means that the corresponding transformation of projective n-space also has a fixed point.

Every one-valued continuous transformation of the interior and boundary of an $(n-1)$-sphere into itself or a part of itself has an invariant point.

For we may assume that the $(n-1)$-sphere is situated in n-space of inversion which is homeomorphic with an n-sphere. It is then possible to extend the definition of the transformation to the entire n-space by specifying that every point exterior to the $(n-1)$-sphere shall be carried into the same point as its reflection within the sphere. Now, the transformation thus defined carries the entire n-space into a set of points bounded by the $(n-1)$-sphere. It is therefore of index 0 and consequently has an invariant point. Obviously, the fixed point is on or interior to the $(n-1)$-sphere, since all exterior points are carried into interior points or points on the $(n-1)$-sphere.

On an n-sphere of even dimensions, there cannot exist a one-valued vector distribution which is everywhere continuous.

For if there existed such a distribution, the points P of the n-sphere could all be displaced by a constant amount $\delta < 2\pi$ along great circles tangent to the vectors at the points, and the n-sphere would therefore undergo a transformation into itself without fixed point. But the index of this transformation would clearly be $+1$, since we could pass continuously from S_1 to S_2 by allowing all the points P to travel at a uniform rate along their respective great circles. We would thus be led to a contradiction with the theorem of §5.

7. The extended form of Brouwer's theorem also lends itself to a number of new applications. We indicate one of them by way of illustration.

Let R be the interior and boundary of a finite region of the plane bounded by p non-intersecting circles, one enclosing all the others. We consider a transformation of R into itself which is continuous on R, which has a one-valued inverse on the boundary of R, and which carries each of the bounding circles into itself with preservation of sense. Then the transformation always has an invariant point if p is not equal to 2.

For, here again, we may extend the transformation over the entire plane in such a way that the exterior of the outer circle is carried into its interior and the interiors of the $p-1$ inner circles into their exteriors. We then obtain a transformation of the plane of inversion which is of index $1-p$. Consequently,

the transformation has a fixed point if p is not 2. Clearly, also, the fixed point is a point of R.

It would be easy to generalize the last theorem to n dimensions or to examine the case where the bounding circles are permuted, transformed into themselves with given indices, and so on. However, it seems hardly necessary to dwell on such obvious extensions.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

INVARIANT POINTS IN FUNCTION SPACE*

BY

G. D. BIRKHOFF AND O. D. KELLOGG

Existence theorems in analysis deal with functional transformations. This suggests that such existence theorems may be obtained from known theorems on point transformations in space of two or of three dimensions by generalization, first to space of n dimensions, and then to function space by a limiting process. This direction of attack has been followed out and has resulted in the theorems given below. For instance it is found that theorems on invariant points for the sphere or for its surface yield respectively by generalization existence theorems in analysis of non-homogeneous and of homogeneous type.

The treatment is here confined to the case of real functions of a real variable, although extensions to real functions of several real variables are indicated. Only the case of a single unknown function is considered. In many cases, of course, apparently more general problems can be reduced to this case by a process which is familiar in the theory of integral equations, namely the juxtaposition of intervals.

The applications include the classical existence theorems for differential and integral equations, linear and non-linear.

Incidentally, it is proved that an algebraic manifold $f_1 = c_1, f_2 = c_2, \ldots, f_m = c_m$, where $f_1, f_2, \ldots, f_m$ are real polynomials in the real variables $x_1, x_2, \ldots, x_n$, has no singularity for general values of the real constants $c_1, c_2, \ldots, c_m$. The authors have not been able to find any earlier proof of this simple and important theorem.

The literature on the subject of invariant points does not appear to be extensive. For a geometric treatment of one-valued transformations with one-valued inverses, we may refer to L. E. J. Brouwer.† Some existence theorems of un-

* Presented to the Society, Dec. 30, 1920 and Feb. 25, 1922.

† *Ueber eindeutige stetige Transformationen von Flächen in sich*, Mathematische Annalen, vol. 69 (1910), p. 176. See his references there to the Proceedings of the Section of Sciences of the Royal Academy at Amsterdam, and in addition, ibid., vol. 13 (1911), pp. 771, 777.

Since the writing of this paper (see, however, footnote on p. 111), we have learned from Professor J. W. Alexander that he has obtained results on the existence of invariant points, which, had we known of them, would have sufficed as the basis in n-space for a considerable portion of our theory. The importance of the subject, however, and the difference in method appear to us to warrant the inclusion here of our proofs of the needed theorems. Professor Alexander's paper will be found in these Transactions, vol. 23, 1922, pp. 89–95.

usual generality have been developed previously. Thus Mason has given a simple theorem for the linear non-homogeneous problem.* Evans has treated certain extensions of the integral equation of Volterra type.† Mrs. Pell‡ has considered linear transformations of general type.

1. A LEMMA ON ALGEBRAIC EQUATIONS

Our proofs of the theorems on the existence of invariant points in n-space will be based on the method of analytic continuation. In this paragraph we shall establish a lemma affirming the validity of this method under certain conditions by means of the following theorem, a corollary of which is that the general algebraic manifold is non-singular.

THEOREM. *Let $f_1, f_2, \ldots, f_m$ be m polynomials in the n variables $x_1, x_2, \ldots, x_n$, where $m < n$. If c_1 is chosen not to have one of a finite number of values, after which c_2 is chosen not to have one of a finite number of values, and so on, until c_m is chosen, then the equations $f_1 = c_1, f_2 = c_2, \ldots, f_m = c_m$ will have no solution for which the rank of the m-rowed matrix*

$$M \equiv \left\| \begin{array}{cccc} \dfrac{\partial f_1}{\partial x_1}, & \dfrac{\partial f_1}{\partial x_2}, & \ldots, & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1}, & \dfrac{\partial f_2}{\partial x_2}, & \ldots, & \dfrac{\partial f_2}{\partial x_n} \\ \ldots & \ldots & \ldots & \ldots \\ \dfrac{\partial f_m}{\partial x_1}, & \dfrac{\partial f_m}{\partial x_2}, & \ldots, & \dfrac{\partial f_m}{\partial x_n} \end{array} \right\|$$

is less than m.

This may be proved by induction. Let us first consider the case $m = 1$. It is known that any set of algebraic equations is satisfied at a finite number of isolated points and on a finite number of analytic manifolds of various dimensions.§ On any manifold for which all the partial derivatives of f_1 vanish, it is clear that $df_1 = 0$, and hence $f_1 = \text{const}$. The value of c_1 will be

---

* *Selected topics in the theory of boundary value problems of differential equations, New Haven Colloquium Lectures*, Yale University Press, 1910, pp. 174 ff.

† Proceedings of the Fifth International Congress of Mathematicians at Cambridge (1912), p. 387. See also *Functionals and their applications, Cambridge Colloquium Lectures*, New York, 1918.

‡ *Biorthogonal systems of functions*, these Transactions, vol. 12 (1911), p. 135, and *Applications of biorthogonal systems of functions to the theory of integral equations*, ibid., p. 165.

§ The facts are due to Weierstrass. Complete proofs of them will be given in the second volume of Osgood's *Funktionentheorie*. They can also be established by means of Kronecker's theory of elimination. See his *Grundzüge einer arithmetischen Theorie der algebraischen Grössen*, Journal für Mathematik, vol. 92 (1889), p. 28 ff.

taken distinct from any of the finite number of values of f_1 on these manifolds or at the isolated points referred to. Thus the surface $f_1 = c_1$ will be non-singular, and the theorem holds for $m=1$.

We now assume the theorem for $m-1$ polynomials. Let the values of c_1, $c_2, \ldots, c_{m-1}$ be so chosen that the equations

$$f_1 = c_1, f_2 = c_2, \ldots, f_{m-1} = c_{m-1} \tag{1}$$

define non-singular manifolds, i. e., such that th ematrix M' obtained from M by omitting the last row is of rank $m-1$ at all points of the manifolds.

Consider the locus defined by the equations (1) together with those obtained by equating to 0 all the determinants of order m of the matrix M. As before, this locus consists in a finite number of isolated points and of analytic manifolds On any such manifold, $f_m = \text{const}$. For, differentiating (1), we have

$$\begin{aligned}
df_1 &= \frac{\partial f_1}{\partial x_1}\,dx_1 + \frac{\partial f_1}{\partial x_2}\,dx_2 + \cdots + \frac{\partial f_1}{\partial x_n}\,dx_n = 0,\\
df_2 &= \frac{\partial f_2}{\partial x_1}\,dx_1 + \frac{\partial f_2}{\partial x_2}\,dx_2 + \cdots + \frac{\partial f_2}{\partial x_n}\,dx_n = 0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
df_{m-1} &= \frac{\partial f_{m-1}}{\partial x_1}\,dx_1 + \frac{\partial f_{m-1}}{\partial x_2}\,dx_2 + \cdots + \frac{\partial f_{m-1}}{\partial x_n}\,dx_n = 0,
\end{aligned}$$

and since the matrices M and M' are both of rank $m-1$, these equations have as consequence

$$df_m = \frac{\partial f_m}{\partial x_1}\,dx_1 + \frac{\partial f_m}{\partial x_2}\,dx_2 + \cdots + \frac{\partial f_m}{\partial x_n}\,dx_n = 0,$$

or, $f_m = \text{const}$. If, therefore, the constant c_m is distinct from the values of f_m at the isolated points and on those manifolds for which the equations (1) hold and all the determinants of order m of the matrix M vanish, it will follow that the locus $f_1 = c_1, f_2 = c_2, \ldots, f_m = c_m$ is non-singular, as was to be proved.

Of course the precise conditions for singularities consist in certain algebraic relations between $c_1, c_2, \ldots, c_m$. In the non-singular case the manifolds are of dimensions $n-m$. It is not difficult to extend the theorem to the case where the functions f_i are merely restricted to be analytic.

We are now in a position to prove the desired lemma.

LEMMA. *Let $G_1, G_2, \ldots, G_{n-1}$ denote $n-1$ polynomials with real coefficients in the n variables $x_1, x_2, \ldots, x_n$, and let C denote a bounded open continuum of the real space of these variables. If there exists a point A on the boundary of C at which the polynomials G_i all vanish, while one of their functional determinants with respect to $n-1$ of the variables x_i does not vanish, and if there exists in every*

neighborhood of A both points in C and points without C at which the G_i all vanish, then there exists a point B on the boundary of C, distinct from A, at which they all vanish.

From the theorem just proved, we know that, given any positive ϵ, there exist sets c_i less in absolute value than ϵ such that any curve, K_c, satisfying the $n-1$ equations in $x_1, x_2, \ldots, x_n$

$$G_1 = c_1, G_2 = c_2, \ldots, G_{n-1} = c_{n-1} \tag{2}$$

is non-singular. But for $c_1 = c_2 = \cdots c_{n-1} = 0$ there is, according to the theory of implicit functions, a curve branch K_0 through A satisfying the corresponding equations (2) and passing into C.

As the constants c_i are varied, the coördinates and direction cosines of K_c vary analytically. It is possible to describe a small sphere S about A, which K_0 will cut twice, and but twice, one of these times in a point interior to C. The number ϵ can be taken so small that for all c_i, less in absolute value than ϵ, the curve K_c will, on the one hand, cut the sphere twice and only twice, one of these times in the interior of C, and, on the other hand, will contain points exterior to C and to S.

Now consider an infinite sequence of these curves, corresponding to sets of values of the c_i, for each of which the curves K_c defined by (2) are non-singular, while $\lim c_i = 0$. Each such curve can be continued analytically until it leaves C at some point B' outside of the sphere S. The set of points B' corresponding to the infinite sequence of curves K_c is bounded, and hence must have at least one limit point B, outside of S. Such a point B is a boundary point of C because the points B' are boundary points of C. Finally, since the functions G_i are continuous, the functions G_i vanish at B. Thus the lemma is established.

2. Invariant points in n-space

We proceed to the proof of the following theorem:

Theorem I. *Let R_n denote a bounded connected region of real n-space containing an interior point O (the origin for a set of rectangular coördinates $x_1, x_2, \ldots, x_n$) such that every half-ray originating in O contains but one boundary point of R. Let T denote a one-valued and continuous transformation*

$$x'_1 = f_1(x_1, x_2, \ldots, x_n),\ x'_2 = f_2(x_1, x_2, \ldots, x_n), \ldots,$$
$$x'_n = f_n(x_1, x_2, \ldots, x_n), \text{ or briefly, } x' = f(x),^* \tag{3}$$

which transforms each point of R_n into a point of R_n.

* A symbol without index will, in the following, be understood as standing for the totality of the corresponding symbols with indices.

Then there exists a point, a, of R_n, which is invariant under T, i. e., such that $a = f(a)$.*

We will prove the theorem first under the assumption that the functions $f(x)$ are polynomials, and will then pass to the general case by means of Weierstrass's theorem on the approximation to continuous functions by polynomials.

Consider the transformation T_λ

$$x' = \lambda f(x) \tag{4}$$

in which the parameter λ lies in the closed interval (0, 1). For $\lambda = 0$, this transformation has the single invariant point $x = 0$. Our object is to show that it has an invariant point for $\lambda = 1$. We are thus led to study the set of n algebraic equations

$$x' - \lambda f(x) = 0 \tag{5}$$

in the cylindrical region $R_{n\lambda}$ of $(n + 1)$-space: x in R_n and λ in (0, 1).

For $x = 0$, $\lambda = 0$, the functional determinant of the functions $x - \lambda f(x)$, with respect to the n variables x, is 1. Hence, in the neighborhood of this point, the values x are determined by (5) as real analytic functions of λ, and, for small negative and positive values of λ, these equations have solutions without and within $R_{n\lambda}$, respectively. The hypotheses of our lemma are thus satisfied, and there exists a second point, B, on the boundary of $R_{\lambda n}$, whose coördinates satisfy (5). For this point λ is not 0, since $x = \lambda = 0$ is the only point on $\lambda = 0$ whose coördinates satisfy the equations (5). Furthermore B does not lie on the boundary of $R_{n\lambda}$ for $0 < \lambda < 1$, since T_λ carries all such points into the *interior* of R_n. Hence B must lie on the boundary $\lambda = 1$; and, if $(a, 1)$ denote the coördinates of B, we have $a = f(a)$, as was to be proved.

Suppose the functions $f(x)$ are not polynomials. We shall find it convenient to develop first a property of the region R_n. Let α stand for the direction cosines of a half-ray originating in O, and let $r(\alpha)$ denote the distance from O to the single boundary point of R_n on this half-ray. Then $r(\alpha)$ is a continuous function of the variables α. For, if discontinuous, say at a point α_0, there would exist a positive number η, and an infinite sequence $[\alpha_i]$ of directions with α_0 as limit, such that $|r(\alpha_i) - r(\alpha_0)| \geqq \eta$. Since R_n is bounded, the boundary points corresponding to $[\alpha_i]$ would have at least one limit point on the half-ray α_0 with distance from O different from $r(\alpha_0)$. Such a limit point would be a boundary point of R_n, a conclusion in contradiction with the hypothesis that each half-

* The theorem is stated and proved in a degree of generality sufficient for our later purposes. It will be seen that the proof holds if merely the boundary points of R_n are transformed into points of R_n, and the theorem admits the obvious extension to any region susceptible of continuous one to one mapping on a region R_n of the kind described. See also Alexander, loc. cit.

ray from O contains but one boundary point of R_n. Hence $r(\alpha)$ is continuous, and, because of the closure of the domain of the variables α, uniformly continuous. There exists, therefore, a function, $\varphi(\Delta\theta)$, approaching 0 with $\Delta\theta$, but positive for $\Delta\theta > 0$, such that $|r(\alpha') - r(\alpha)| \leqq \varphi(\Delta\theta)$ for any two directions α and α' making an angle $\Delta\theta$ or less with each other. This function may be so chosen as to exceed in magnitude the corresponding chord also. This is the needed property of R_n.

Now for any small positive ϵ there exist n real polynomials, $p(x)$, such that

$$|p_i(x) - f_i(x)| \leqq \epsilon \tag{6}$$

for $i = 1, 2, \ldots, n$ and for all x in R_n. The transformation $\Pi: x' = p(x)$ would be of the type for which we have proved the existence of an invariant point, if it transformed points of R_n into points of R_n. In this event the theorem would follow immediately inasmuch as a limit point of the invariant points under Π as ϵ approaches 0 is clearly an invariant point under T.

In case Π does not transform all points of R_n into points of R_n, we shall modify Π slightly to a suitable new polynomial transformation $\Pi_k: x' = kp(x)$ where k is slightly less than 1. Let A be any point of R_n, and B the point into which the transformation $x' = p(x)$ carries A. We are only interested in the possible case that B is not in R_n, since otherwise the point B_k given by $x' = kp(x)$ lies in R_n also. On this assumption, let C be the nearest point of R_n to B. Then $BC \leqq \sqrt{n}\epsilon$, because T carries A into a point of R_n and because of the inequalities (6). If m denote the positive minimum of $r(\alpha)$, the angle subtended by BC at O is not greater than $2 \sin^{-1}(\sqrt{n}\epsilon/2m)$. Then, if D is the boundary point of R_n on the half-ray OB, $CD \leqq \phi(2 \sin^{-1}\sqrt{n}\epsilon/2m)$. Hence

$$BD \leqq BC + CD \leqq \sqrt{n}\epsilon + \phi(2 \sin^{-1}(\sqrt{n}\epsilon/2m)),$$

and thus there is a small upper limit for the distance from B to the nearest point D of R_n on the half-ray OB.

If now k is taken not greater than the least value of OD/OB, all points B will be brought within R_n by Π_k. If the upper limit for BD be denoted by δ, we have

$$OD/OB \leqq OD/(OD + \delta) = r(\alpha)/(r(\alpha) + \delta)$$

which is least when $r(\alpha)$ has its least value, m. Hence if $k = m/(m + \delta)$, the transformation Π_k will carry each point of R_n into a point of R_n.

The degree of approximation of Π_k to T follows from the identity

$$f_i(x) - kp_i(x) \equiv k[f_i(x) - p_i(x)] + (1 - k)f_i(x)$$

whence, if M is the maximum of $r(\alpha)$,

$$|f_i(x) - kp_i(x)| \leqq \epsilon + (1 - k)M. \tag{7}$$

Since k approaches 1 as ϵ approaches 0, the approximation may be made arbitrarily close.

Thus we infer as before the existence of an invariant point under T as a limit of the invariant points of Π_k for $\lim \epsilon = 0$.

3. The transition to function space; first method

The transition from n-space to function space may be made in a variety of ways. First we give one based on interpolation; in §5 we shall give one based on the mediation of a Hilbert space.

It will be convenient to define a few terms in advance. We shall be concerned with real functions, $f(s)$, defined on some interval, say (0, 1). A set of such functions will be *bounded*, B, provided a constant B exists such that $|f| \leqq B$ for every function of the set, and for every s in (0, 1). The set will be said to be *equicontinuous*, $\eta(\epsilon)$, provided there exists a function, $\eta(\epsilon)$, defined and bounded for $0 \leqq \epsilon \leqq 1$, and approaching 0 with ϵ, such that $|f(s+h) - f(s)| \leqq \eta(\epsilon)$, whenever $|h| \leqq \epsilon$, and s and $s + h$ are in (0, 1), for every function of the set.* A single function of such a set will be said to be *continuous*, $\eta(\epsilon)$. An infinite sequence of bounded equicontinuous functions has the fundamental property that a subsequence can be so chosen as to approach a function of the set uniformly.

A function $\eta(\epsilon)$ will be said to be *convex*, provided that, for every a, b, and θ in (0, 1),

$$\eta(a + \theta(b - a)) \geqq \eta(a) + \theta(\eta(b) - \eta(a)).$$

If a set of functions is equicontinuous, $\xi(\epsilon)$, there exists a convex function, $\eta(\epsilon) \geqq \xi(\epsilon)$, approaching 0 with ϵ. Of course the same set of functions is equicontinuous, $\eta(\epsilon)$.

A primary property of functions continuous, $\eta(\epsilon)$, $\eta(\epsilon)$ convex, is that any function $g(s)$, which meets the continuity requirement for $s + h$, s on the set $s_1 = 0, s_2, \ldots, s_n = 1$ and which is *linear* between these values of s, will be continuous, $\eta(\epsilon)$, throughout.

This is self-evident when $\eta(\epsilon) = c\epsilon$, when the continuity requirement is merely that every chord of the given curve $y = f(s)$ has a slope which does not exceed c; for this is true of the polygonal curve $y = g(s)$.

Moreover, in any case, we have by hypothesis,

$$|g(s_j) - g(s_i)| \leqq \eta(s_j - s_i)$$

* Cf. Ascoli, *Le curve limite di una varietà data di curve*, Atti della Reale Accademia dei Lincei, Memorie, ser. 3, vol. 18 (1882–83), pp. 545–549.

for $s_j > s_i$, say. Since the curve $y = \eta(s - s_i)$ is convex in the positive y direction for $s > s_i$, it follows that for any $s > s_i$

$$| g(s) - g(s_i) | \leqq \eta(s - s_i).$$

If now $s_i \leqq s' \leqq s_{i+1} \leqq s$, and we write

$$\lambda = \frac{s' - s_i}{s_{i+1} - s_i}, \quad \mu = \frac{s_{i+1} - s'}{s_{i+1} - s_i},$$

so that $\lambda + \mu = 1$, the polygonal character of $g(s)$ yields

$$| g(s) - g(s') | = | \lambda(g(s) - g(s_{i+1})) + \mu(g(s) - g(s_i)) | \leqq \lambda\eta(s - s_{i+1}) + \mu\eta(s - s_i).$$

But the right hand side of this last inequality represents the ordinate of the chord of $y = \eta(x)$ at $s - s'$, with end points at $s - s_{i+1}$ and $s - s_i$, so that

$$| g(s) - g(s') | \leqq \eta(s - s').$$

This same inequality obviously holds for s, s' lying on one and the same interval (s_i, s_{i+1}). Hence $g(s)$ is continuous, $\eta(\epsilon)$.

It may be observed that the proofs of the paper become somewhat simpler for the important case $\eta(\epsilon) = c\epsilon$, referred to above.

A transformation $f' = Sf$ will be said to be *continuous* in a region R_f of function space provided for every $\epsilon > 0$, and for every function $g(s)$ in R_f, there exists a positive number δ, such that $| Sf - Sg | \leqq \epsilon$ for all s in (0, 1) whenever $| f - g | \leqq \delta$, uniformly with respect to s, f being in R_f.

THEOREM II. *Let R_f denote the totality of real functions $f(s)$, defined on the closed interval (0, 1), which are bounded, B, and equicontinuous, $\eta(\epsilon)$, $\eta(\epsilon)$ being convex. Let $f' = Sf$ denote a one-valued, continuous transformation which carries each point of R_f into a point of R_f.*

Then there exists a point of R_f which is invariant under this transformation.

The "distance," $\delta_f = \sqrt{\int_0^1 (f(s) - Sf(s))^2 ds}$, by which a point f is moved by the transformation S, has a lower limit $\delta_0 \geqq 0$ in R_f. This limit is attained at some point, say f_0, of R_f. For if $[f_i]$ is an infinite sequence of points for which δ_f approaches δ_0, there exists a subsequence which approaches uniformly f_0 in R_f so that $\delta_{f_0} = \delta_0$.

If $\delta_0 = 0$, it is clear that f_0 is an invariant point and the theorem follows.

To show that $\delta_0 = 0$ we choose a positive n and define a region R_n of n-space by the inequalities

$$| x_i | \leqq B, \ | x_{i+j} - x_i | \leqq \eta(j/(n-1)) \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, n - i).$$

This n-dimensional region satisfies the requirements of Theorem I, being convex toward the interior point O. Next we define the transformation, T, of R_n, as follows. For x in R_n, we construct the continuous polygonal function of s, $\pi(s, x)$, equal to x_i for $s = s_i = (i-1)/(n-1)$, $(i = 1, 2, \ldots, n)$, and linear for intermediate values of S. This function is in R_f, for it is evidently bounded, B, and is continuous, $\eta(\epsilon)$, as has been seen (p. 102.)

Let $\pi'(s, x)$ denote the function of R_f into which $\pi(s, x)$ is transformed by S.* Then, if x' be defined by the equations $x'_i = \pi'(s_i, x)$, the transformation T is determined. This transformation is easily seen to have the properties required in Theorem I, and so has an invariant point, a, in R_n.

The function $\pi(s, a)$, formed for this point, is transformed by S into a function $\pi'(s, a)$ which coincides with it at the n points s_i. Between these points, the variation of either function is not greater than $\eta(1/(n-1))$. Hence, throughout the interval $(0, 1)$, $|\pi(s, a) - S\pi(s, a)| \leqq 2\eta(1/(n-1))$. Consequently, $\delta_\pi \leqq 2\eta(1/(n-1))$, a quantity which approaches 0 as n increases indefinitely; and $\delta_0 = 0$, as was to be proved.

4. Extensions

It is, of course, evident that any other finite interval than $(0, 1)$ might have been used. Even infinite intervals may be used if the equicontinuity hypothesis holds for some new variable, such as $1/s$.

Moreover, the reasoning requires no essential modification in order to be applicable to functions of two or more variables. For instance a set of functions $f(s, t)$ defined on the square $0 \leqq s, t \leqq 1$ can be defined as equicontinuous, $\eta(\epsilon)$, if equicontinuous, $\eta(\epsilon)$, in each variable separately. We should compare the function space thus defined with a space of n^2 dimensions. To the point $f(s, t)$ of function space, we can make correspond the point x of n^2-space defined by coördinates $x_{i,j} = f((i-1)/(n-1), (j-1)/(n-1))$, the subscripts of x corresponding to a network of vertices of squares of sides $1/(n-1)$ in the original square $0 \leqq s, t \leqq 1$. Given a point in this n^2-space, the intermediate coordinates of points of function space can be defined as the value of the bilinear function $ast + bs + ct + d$ which has the same values at the vertices of the corresponding square.

Another definition of equicontinuity, $\eta(\epsilon)$, of functions $f(s,t)$ might be employed namely that $|f(s_2, t_2) - f(s_1, t_1)| \leqq \eta(\epsilon)$ for $\sqrt{(s_2 - s_1)^2 + (t_2 - t_1)^2} \leqq \epsilon$. It is readily seen that this type of equicontinuity involves the preceding, and that conversely if the preceding type holds so will this second type, provided that $\eta(\epsilon)$ is replaced by $2\eta(\epsilon)$.

* To avoid possible confusion attention is called to the fact that the prime is used to denote the result of the given transformation, S.

Theorem II may, in fact, be regarded as a type of a large body of theorems that can be established. The region R_f may be modified in a variety of ways. Important for the present method of proof are the closure and convexity of R_f, and the continuity of T. One direction of modification of the theorem is, however, of sufficient importance to warrant more detailed attention. We formulate it in the following theorem.

THEOREM III. *Let R_f denote the totality of real functions f defined on $(0, 1)$, which are bounded, B_0, and whose first derivatives exist and are bounded, B_1, and are equicontinuous $\eta(\epsilon)$, $\eta(\epsilon)$ being convex. Let $f' = Sf$ denote a transformation which is one-valued, carries each point of R_f into a point of R_f, and which is continuous in such wise that, given any function, g, of R_f, and any positive number ϵ, there exists a positive number δ such that $|Sf - Sg| \leqq \epsilon$ for all s, whenever f is any function of R_f for which $|f - g| \leqq \delta$ and $|df/ds - dg/ds| \leqq \delta$ for all s.*

Then there exists an invariant point under S in R_f.

It will be noted that the hypothesis on the continuity of S is weaker than in Theorem II, while the region R_f is more restricted in the present theorem. Theorem III is given with applications to differential equations in view.

The kernel of the proof of the theorem lies, as before, in the choice of a suitable region R_{n+1}, and a transformation, T.

In this case we shall define R_{n+1} by the inequalities

$$|x_i| \leqq B_1, \quad |x_{i+j} - x_i| \leqq \eta(j/(n-1)) \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, n - i),$$

$$|x_0| \leqq B_0, \quad \left|x_0 + \sum_{i=1}^{k} \frac{x_{i+1} + x_i}{2(n-1)}\right| \leqq B_0 \qquad (k = 1, \ldots, n-1).$$

This $(n + 1)$-dimensional region meets the requirements of Theorem I with O as an interior point.

An appropriate determination of T may be made as follows. Given x in R_{n+1}, as defined in the above inequalities, we form the polygonal function $\pi(s, x)$, assuming the values x_i for $s = s_i = (i - 1)/(n - 1)$. We then write

$$\psi(s, x) = l(x_0 + \int_0^s \pi(s, x)ds) \qquad (0 < l < 1).$$

The absolute value of the ordinate of the curve $y = \psi(s, x)$ for $s = s_k$ is l times the left hand member of the last inequality defining R_{n+1}. Hence it is clear that $\psi(s, x)$, for $s = s_k$, does not exceed lB_0 in numerical value. Since $|d\psi/ds| \leqq lB_1$, we infer that

$$|\psi(s, x)| \leqq l(B_0 + B_1/(n - 1)) \leqq B_0$$

for n large, while $d\psi/ds$ is bounded, B_1, and continuous, $\eta(\epsilon)$. Thus $\psi(s, x)$

lies in R_f, and consequently so does $S\psi(s, x)$. The definition of T is then given by the equations

$$x_0' = \psi'(0, x), \quad x_i' = \frac{d}{ds}\,\psi'(s_i, x) \qquad (i = 1, 2, \ldots, n,)$$

where

$$\psi'(s, x) = l\,S\psi(s, x).$$

It is immediately evident that T is a one-valued continuous transformation throughout R_{n+1} because of the corresponding restriction on S. In order to prove that T has an invariant point, we need only show further that the point x' lies in R_{n+1}. Since ψ' is bounded, lB_0, we have

$$\left|\psi'(0, x) + \int_0^s \frac{d\psi'}{ds}\,ds\right| \leqq l\,B_0,$$

whence, for $s = s_k$, we infer

$$\left|x_0' + \sum_{i=1}^{k} \frac{x_i' + x_{i+1}'}{2(n-1)}\right| \leqq l\left(B_0 + \eta\left(\frac{1}{n-1}\right)\right)$$

inasmuch as $d\psi'/ds$ is continuous, $l\eta(\epsilon)$. Consequently, for any fixed positive l, n may be chosen so large that the last inequality defining R_{n+1} is satisfied by x'. And since $d\psi'/ds$ is bounded, B_1, and continuous, $\eta(\epsilon)$, it is clear at once that the other requirements are satisfied by x'. Hence an invariant point, a, exists for any such l, if n is large enough.

For the function $\psi(s, a)$ of R_f corresponding to this invariant point, it follows from the invariance of a under T that

$$l\,\frac{d}{ds}\,S\psi(s, a) = \frac{d}{ds}\,\psi(s, a), \text{ for } s = s_i \qquad (i = 1, 2, \ldots, n),$$

$$l\,S\psi(0, a) = \psi(0, a).$$

From the first of these equations results the inequality

$$\left|\frac{d}{ds}S\psi(s, a) - \frac{d}{ds}\,\psi(s, a)\right| \leqq (1 - l)B_1$$

for the same values of s, so that for any s the left hand member is less than

$$(1 - l)\,B_1 + \eta(1/(n - 1)).$$

From the second equation follows

$$|\,S\psi(0, a) - \psi(0, a)\,| \leqq (1 - l)B_0.$$

Hence

$$|\,S\psi(s, a) - \psi(s, a)\,| \leqq (1 - l)\,(B_0 + B_1) + \eta\left(\frac{1}{n-1}\right)$$

so that the "distance," δ_ψ, between ψ and $S\psi$ in R_f, can be made arbitrarily small by taking l near enough to 1 and n large enough.

By the argument used in the proof of Theorem II it follows that there exists an invariant function in R_f.

We close this paragraph with a formulation of a generalization of Theorem III which is useful in applications, and which may be proved by a simple extension of the reasoning used to establish that theorem.

THEOREM IV. *Let R_f denote the totality of real functions defined on (0, 1) which, with their first $n-1$ derivatives are bounded, $B_0, B_1, B_2, \ldots, B_{n-1}$, respectively, and whose nth derivatives are bounded, B_n, and equicontinuous, $\eta(\epsilon)$, $\eta(\epsilon)$ being convex. Let $f' = Sf$ denote a transformation which is one-valued, carries each point of R_f into a point of R_f, and is continuous in such wise that given any function g of R_f and any positive number ϵ, there exists a positive number δ such that $|Sf - Sg| \leqq \epsilon$ for all s, whenever f is any function of R_f for which $f-g$ and its first $n-1$ derivatives nowhere exceed δ in absolute value.*

Then there exists an invariant point under S in R_f.

5. THE TRANSITION TO FUNCTION SPACE; A SECOND METHOD

A second method of establishing the existence of an invariant point in a function space consists in setting up a correspondence with a certain Hilbert space, i. e., a space of countably infinitely many dimensions such that the sum of the squares of the coördinates of each point converges. The correspondence is based on a closed set of continuous bounded orthogonal functions $[\phi]$, such as the set appearing in Fourier's series, so that no function with summable square is orthogonal to the set unless it vanishes except on a set of zero measure. We shall denote the generalized Fourier coefficients of a function f by a_i, i. e., $a_i = \int_0^1 f\phi_i \, ds$. We may then state

THEOREM V. *Let R_f denote the totality of summable functions with summable squares, $f(s)$, $(0 \leqq s \leqq 1)$, such that*

(a) there exists a constant B such that $\int_0^1 f^2 ds \leqq B^2$, and

(b) there exists a function $\eta(m)$, approaching 0 with $1/m$, such that $\sum_m^\infty a_i^2 \leqq \eta(m)$. Let $f' = Sf$ denote a transformation which is one-valued, carries each point of R_f into a point of R_f, and is continuous in such wise that given g in R_f and $\epsilon > 0$, there exists a $\delta > 0$ such that $\int_0^1 (Sf - Sg)^2 ds \leqq \epsilon$ for all points f of R_f for which $\int_0^1 (f-g)^2 ds \leqq \delta$.

Then there exists an invariant point under S in R_f, where it is understood that two points are to be regarded as identical provided their coördinates differ at most on a set of zero measure.

We first establish the closure of the space R_f. Given an infinite set $[f]$ of points of R_f, an infinite sequence $[f_j]$, can be selected from this set which converges in the mean to a function f of R_f. To show this, we consider the first Fourier coefficients of the set $[f_j]$. As these first coefficients are bounded, B, and infinitely numerous, an infinite sequence can be selected from the set $[f_j]$, whose first Fourier coefficients approach a limit, a_1'. From this sequence can be selected an infinite sub-sequence whose second Fourier coefficients approach a limit a_2'. And from this subsequence, in turn, another, whose third Fourier coefficients approach a limit a_3', and so on.

These constants $(a_1', a_2', a_3', \ldots)$, are the Fourier constants of a function f of R_f. For, the sum of the squares of the constants a_i' converges to a limit not exceeding B^2. But this, together with the properties assumed for the set $[\phi_i]$ at the outset, is sufficient, according to the Riesz-Fischer theorem,* for the existence of a summable function with summable square, whose Fourier coefficients are the a_i'. This is the function f whose existence was asserted, for it is easily verified that the hypotheses (a) and (b) are fulfilled.

If now we select the first function of the first sequence, the second function of the second sequence, and so on, we have a single sequence $[f_j]$ which converges in the mean to f, as may be seen by use of hypothesis (b) and the equation $\int_0^1 (f_j - f)^2 ds = \sum_1^\infty (a_{ji} - a_i')^2$. Thus the region R_f is closed in the sense that every infinite set of points in it contains a sequence which converges in the mean to a point of R_f.

It is a consequence of this closure, and of the continuity of the transformation, that the distance $\delta_f = \sqrt{\int_0^1 (Sf - f)^2 ds}$, by which the transformation S displaces the point f, attains its lower limit δ_0 in R_f. That this lower limit cannot be different from 0 is seen by a comparison with a region R_n of n-space, namely the region defined by $\sum_1^n x_i^2 \leqq B^2$, $\sum_m^n x_i^2 \leqq \eta(m)$, $(m = 1, 2, \ldots, n)$. To a point x of this region corresponds the point $\phi(s, x) = \sum_1^n x_i \phi_i$ of R_f. Let $\phi'(s, x)$ denote the function into which S transforms $\phi(s, x)$. Then the transformation T is defined by the equations $x_i' = \int_0^1 \phi'(s, x)\phi_i(s) ds$, $(i = 1, 2, \ldots, n)$. The

* Cf. E. Fischer, *Sur le convergence en moyenne*, Paris Comptes Rendus, vol. 144 (1907), pp. 1022–24.

hypotheses of Theorem I are fulfilled, and an invariant point, a, exists. Moreover, δ_ϕ approaches 0 with $1/n$, so that $\delta_0 = 0$. But the point for which $\delta_0 = 0$ is an invariant point, in the specified sense.

6. Applications

There are evidently numerous applications of the above theorems to the existence problems of analysis. We give two instances in this paragraph.

In the first place, let us consider a differential equation

$$y^{(n)} = F(x, y, y', \ldots, y^{(n-1)}) \tag{8}$$

for which a solution is desired which satisfies n linear conditions on the interval $(0, a)$

$$\int_0^a \sum_{j=0}^{n-1} p_{ij}(x) y^{(j)}(x)\,dx + \sum_{j=0}^{n-1}\sum_{k=1}^{m} q_{ijk} y^{(j)}(x_k) = c_i \qquad (i = 1, 2, \ldots, n; 0 \leqq x_1 \leqq x_2, \ldots, x_m \leqq a), \tag{9}$$

where the functions $p_{ij}(x)$ are continuous and the conditions are such as to determine uniquely a polynomial y of degree $n-1$. The problem of proving the existence of a solution of the differential equation (8) with auxiliary conditions (9) is identical with the problem of proving the existence of an invariant point of the transformation $z = Sy$;

$$z = \int_0^x \int_0^x \cdots \int_0^x F(x, y, y', \ldots, y^{(n-1)})\,dx\,dx \ldots dx + a_0 + a_1 x + \cdots + a_{n-1} x^{n-1},$$

the coefficients a_i being explicit functionals of y determined by the demand that z satisfy the conditions (9). But here Theorem IV gives information, so that we may state the following corollary:

If there exists a set of n constants, $B_0, B_1, \ldots, B_{n-1}$, such that when $|y| \leqq B_0$, $|y'| \leqq B_1, \ldots, |y^{(n-1)}| \leqq B_{n-1}$, the same inequalities hold for z and its derivatives, and if F is a one-valued and continuous function of its arguments thus restricted and with $0 \leqq x \leqq a$, then there exists in R_f a solution of the differential equation and the auxiliary conditions.

The region R_f will be determined by the given inequalities on $y, y', \ldots, y^{(n-1)}$ together with the requirement that $y^{(n-1)}$ is continuous, $\eta(\epsilon) = M\epsilon$, where M is the maximum numerical value of F for $y, \ldots, y^{(n-1)}$ restricted as in the theorem. It is apparent that $z^{(n-1)}$ is also continuous, $\eta(\epsilon)$.

For a sufficiently small value of a, the conditions stated will be satisfied, so that the above includes the classical existence theorem for the solution of a differential equation with initial values assigned to this solution and its first $n-1$ derivatives at a single point.

Secondly, let us consider an integral equation:*

$$f(s) = y(s) - \lambda \int_0^1 F(y(t))dt \tag{10}$$

in which $f(s)$ is known, and $F(y(t))$ is a given functional of $y(t)$, which may also depend on s and t. We suppose for simplicity that F does not contain derivatives of $y(t)$. If, in this integral equation, $y(s)$ be replaced, outside of the integral sign, by $z(s)$, the equation defines a transformation $z = Sy$, an invariant point of which is a solution of the integral equation. We may therefore conclude from Theorem II, *that, if for continuous functions* $y(t)$ *such that* $|y(t)| \leqq B$, *B greater than the maximum of* $|f(s)|$, *the set of functions of* s, $\int_0^1 F(y(t))dt$, *is bounded, B, and equicontinuous,* $\eta(\epsilon)$, *and if* $f(s)$ *is continuous,* $\eta(\epsilon)$, *then, for sufficiently small values of* λ, *the integral equation has a continuous solution,* $y(s)$. Various extensions will occur to the reader.

7. Invariant directions in a space of an odd number of dimensions

In the above theorems and applications, the invariant point may, in certain cases, turn out to be $f = 0$. For homogeneous problems, however, this would yield only a trivial solution. We therefore now turn our attention to *invariant directions*, in which there is no essential difference between f and kf for $k \neq 0$.

We begin with the case of n-space. It will be convenient to think of the transformations of directions as transformations of points of the hypersphere, H_n : $\sum x_i^2 = 1$, the transformation carrying these points into points other than the origin.

Theorem VI. *Let n be odd, and let T denote a transformation with the following properties:*

(*a*) *it transforms any real point of* H_n *into one real finite point different from the origin;*

(*b*) *it is continuous, i. e., the point x of* H_n *is transformed by it into the point* x' *whose coördinates are continuous functions of the coördinates of x on* H_n.

Then there exists a direction invariant under T on H_n, *i. e., a point x of* H_n *such that* $x' = cx$, $c \neq 0$.

The restriction that n be odd is necessary, since there exist transformations (rotations) in spaces of even dimensionality which have no invariant directions.

Suppose the transformation given in the form $\rho x' = f(x)$, the $f_i(x)$ being continuous, one-valued functions, such that $\rho^2 = \sum f_i^2(x)$ vanishes nowhere on H_n. If the functions $f_i(x)$ are defined only on the hypersphere H_n, their defini-

* See E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen*, Mathematische Annalen, vol. 65 (1908), pp. 370–399.

tion can easily be extended continuously, so that the functions $f_i(x)$ are proportional to the distance $\sqrt{\sum x_i^2}$ of the point x from the origin.

If, then, the transformation have no invariant direction, a sufficiently close polynomial approximation on H_n will have no invariant direction. Assuming, therefore, that the transformation has no invariant direction, we are at liberty to suppose that the $f_i(x)$ are polynomials, and such that $\sum f_i^2(x) > 0$ on H_n. We now form the tangent vector, $t(x) = f(x) - x(\sum x_i f_i(x))$, which never vanishes on H_n since $\sum t_i^2(x) = \sum f_i^2(x) - (\sum x_i f_i(x))^2 = \sum f_i^2(x) \sin^2\theta$, where θ is the angle between the vectors x and $f(x)$. Moreover, $\sum t_i^2(x)$ will remain positive in a finite neighborhood of H_n, and in particular in a closed region S_n bounded by $\sum x_i^2 = r^2$ and H_n, where r is sufficiently near 1.

We then form the vector $s(x) = t(x)\sqrt{\sum x_i^2 / \sum t_i^2(x)}$; it vanishes nowhere in S_n, is perpendicular to the radius x, and its components are analytic throughout S_n. Then $x' = F(x, \lambda) = x\cos\lambda + s(x)\sin\lambda$ is a transformation of S_n into itself, which is analytic for x in S_n and for all λ, and which preserves distances from the origin. As its jacobian, J, is 1 for $\lambda = 0$, the transformation has, for small λ, a one-valued inverse, and the volume S_n may be expressed, for such λ, by the integral

$$V = \int\int \cdots \int_{S_n} J\, dx_1\, dx_2 \ldots dx_n.$$

Hence this integral is constant for small λ, and, being analytic in λ, for all λ. But $J = 1$ for $\lambda = 0$, and $J = -1$ for $\lambda = \pi$. We thus arrive at a contradiction, and an invariant direction must exist.*

8. Invariant directions in function space

From Theorem VI may be derived the following:

Theorem VII. *Let R_f denote the region of function space corresponding to real continuous functions on the interval (0, 1), and let R'_f denote the subregion of R_f corresponding to functions which are bounded, B, and equicontinuous, $\eta(\epsilon)$, $\eta(\epsilon)$ being convex.*

Let $f' = Sf$ denote a transformation applicable to all normalized functions (i. e., such that $\int_0^1 f^2 ds = 1$) of R_f, which yields a unique function in R'_f when so applied,

* After the authors had found proofs of a number of special cases of the above theorem, including the important symmetric case, $f(-x) = -f(x)$ (which includes the linear case), they became acquainted with the details of Professor Alexander's elegant paper, from which the theorem above can readily be inferred. They then completed the general proof above. No other changes in the text have been made since they learned of his results.

is continuous, and is such that the functions f' into which it transforms all normalized functions of R_f have norms, $\int_0^1 f'^2 ds \geqq c^2$, a definite positive number.

Then there exists in R_f' an invariant direction, i. e., a normalized function f such that $f' = kf$, where $k \neq 0$.

We consider the distances

$$\delta_f^{\pm} = \sqrt{\int_0^1 (f \pm kf')^2 ds}$$

between the point f of the hypersphere H_f, $\int f^2 ds = 1$, and the point kf' on H_f corresponding to f by the transformation, and the diametrically opposite point, $-kf'$, respectively. We shall show that the lower limit of either δ_f^+ or δ_f^- must be 0. The existence of an invariant direction will then follow, since the region R_f' is closed.

To show that there are points of H_f in R_f' for which one of these distances is arbitrarily small, we consider an n-space, with n odd, and subject to inequalities to be given presently. We define the transformation T as follows: to x on the hypersphere H_n: $\sum x_i^2 = n-1$, corresponds the continuous function $\pi(s, x)$, equal to x_i for $s = s_i = (i-1)/(n-1)$, $(i = 1, 2, \ldots, n)$, and linear for other values of s. This function is then normalized, giving a function $f(s, x)$. Denoting $Sf(s, x)$ by $f'(s, x)$, we define the point x' on H_n into which T transforms x by the equations

$$x_i' = f'(s_i, x) \Big/ \sqrt{\frac{1}{n-1}\sum f'^2(s_j, x)}.$$

Our first restriction will be to make n so large that T is continuous and Theorem VI applicable. This will be seen to be the case provided the denominator in the expression for x_i' does not vanish. But it is clear that

$$\left| \frac{1}{(n-1)} \sum f'^2(s_i, x) - \int_0^1 f'^2(s, x) ds \right| \leqq 2B\eta \left(\frac{1}{n-1} \right),$$

since the difference in question cannot exceed the maximum variation of $f'^2(s, x)$ in a sub-interval of length $1/(n-1)$, and in turn this variation does not exceed $2B\eta(1/(n-1))$. Inasmuch as $\int_0^1 f'^2(s, x) ds \geqq c^2$, n can be chosen so large that the denominator in the expression for x_i is greater than $c/2$. This is the first condition imposed on n.

Under these circumstances the transformation T has an invariant direction, a, for which $\sum a_i^2 = n-1$. Observing that

$$a_i = \pi(s_i, a) = f'(s_i, a) / \sqrt{\frac{1}{n-1}\sum f'^2(s_j, a)},$$

we conclude that $f(s, a) = c_n Sf(s, a)$ for $s = s_i (i = 1, 2, \ldots, n)$ where $c_n \neq 0$. It follows at once that the δ^* corresponding to the invariant point a in n-dimensional space approaches 0 as n increases, inasmuch as the norm of $Sf \geqq c/2$. This completes the proof of the theorem.

A simple illustration of the Theorem VII is given by the non-linear integral equation

$$\phi(s) = \lambda \int_0^1 K(s, t)\phi^2(t)dt,$$

where $K(s, t)$ is continuous and has a positive lower bound. The right hand side of this equation defines a transformation of functions $\phi(s)$, an invariant direction of which yields a solution of the integral equation. The hypotheses of the theorem are fulfilled, as may be verified by use of the law of the mean, so that a function ϕ exists for which the equation has a continuous normalized solution. A continuous solution, not normalized, evidently exists, then, for every real $\lambda \neq 0$.

9. The existence of inverse directions in n-space

In this, and in the following paragraph, we shall consider the existence of points inverse to a given point for a given transformation. We begin with a transformation in n-space.

Theorem VIII. *Let T_λ denote a transformation $x' = x + \lambda\varphi(x)$, in which the functions $\varphi(x)$ are bounded, B, and continuous, on the hypersphere H_n, $\sum x_i^2 = n-1$. Let λ_0 be a number such that the transformation T_λ transforms no point of H_n into the origin for any value of λ in the closed interval $(0, \lambda_0)$. Let b be any point on H_n.*

Then there exists an inverse point a of H_n, i. e., such that $\rho b = T_{\lambda_0}(a)$, $\rho > 0$.

In other words, any direction has an inverse by T_{λ_0} provided $Q_\lambda : x' = -\lambda\varphi(x)$ has no invariant direction for $0 \leqq \lambda \leqq \lambda_0$.

The proof of the theorem resembles that of Theorem VI and an outline will be sufficient. First assume that the functions $f(x)$ are polynomials, and consider the equations

$$x + \lambda\varphi(x) - \rho b = 0, \qquad \sum x_i^2 = n - 1$$

in the region R_{n+2} : $|x_i| \leqq 2\sqrt{n}$, $\lambda^2 \leqq \lambda_0^2$, $0 \leqq \rho \leqq 1 + B$, B being an upper

bound for $\lambda_0\Phi(x)$ on H_n where $\Phi^2(x) = \sum \phi_i^2(x)/(n-1)$. From the first equations we obtain by squaring and adding

$$\rho^2(n-1) = (n-1) + 2\lambda\sum x_i\varphi_i(x) + \lambda^2\sum \varphi_i^2$$

so that $|\rho| < 1+B$ for any solution $|\lambda| < \lambda_0$. For $\lambda=0$ the equations have a unique solution, for which the functional determinant with respect to the set x and ρ is not 0. The second solution on the boundary of R_{n+2} which by the lemma of §1 exists, necessarily corresponds to the value λ_0 of λ, and a positive value of ρ. If the x have here the values a, then a is the required inverse, and $\rho b = T_{\lambda_0}a$.

If the functions $f(x)$ are merely continuous, the conclusion remains valid, and the proof by a limiting process is immediate.

10. The existence of inverse directions in function space

A theorem in function space which corresponds to Theorem VIII may be formulated as follows.

THEOREM IX. *Let $f' = Sf$ denote a one-valued transformation applicable to continuous functions on H_f and of the form $Sf = f + \lambda Qf$, where $\lambda_0 Qf$ is bounded, B, and equicontinuous, $\eta(\epsilon)$, with $\eta(\epsilon)$ convex. If, in addition, S is continuous, and transforms no function f on H_f into 0 for $0 \leqq \lambda \leqq \lambda_0$ then, for any continuous function g of H_f there exists a direction inverse to g under S_{λ_0} i.e., such that $\rho g = S_{\lambda_0} f$, $\rho > 0$.*

Take any set of variables x_i such that $\sum x_i^2 = n - 1$ and form the corresponding polygonal function $\pi(s, x)$. It is not difficult to establish the inequality

$$\frac{1}{3} \leqq \int_0^1 \pi^2(s, x)dx \leqq \frac{4}{3}.$$

In fact this flows at once from the evident equality

$$\int_{s_i}^{s_{i+1}} \pi^2(s, x)dx = \frac{1}{3(n-1)}(x_i^2 + x_ix_{i+1} + x_{i+1}^2)$$

by summation, since

$$\frac{x_i^2 + x_{i+1}^2}{2} \leqq x_i^2 + x_ix_{i+1} + x_{i+1}^2 \leqq 2(x_i^2 + x_{i+1}^2).$$

Hence, if we normalize $\pi(s, x)$ to $f(s, x)$, the multiplicative factor required lies between $\sqrt{3}$ and $\sqrt{3}/4$, and it evidently approaches 1 as n becomes infinite, provided $\pi(s, x)$, approaches a limit uniformly.

Write now $f' = \pi + \lambda Q f$ and

$$x_i' = f'(s_i, x), \ (s_i = \frac{i-1}{n-1}; i = 1, 2, \ldots, n),$$

thus defining a transformation T_λ of R_n of the form in theorem VIII, namely

$$x_i' = x_i + \lambda Q(f(s_i, x))$$

which is clearly one-valued and continuous. Furthermore, by hypothesis $\lambda Q f$ is bounded, B, and continuous, $\eta(\epsilon)$, for the function $f(s, x)$ on H_f.

Hence by the theorem of §9 we can assert the existence of an inverse point under T unless for $0 \leqq \lambda \leqq \lambda_0$ and some x on H_π,

$$0 = f(s_i, x) + \lambda Q f$$

for $s = s_i$ $(i = 1, 2, \ldots, n)$. Here $\lambda Q f$ is bounded, B, and continuous, $\eta(\epsilon)$, so that $f(s, x)$ is bounded, $B + \eta(1/(n-1))$, and, on account of its polygonal character, continuous, $\eta(\epsilon)$. If such a function $f(s, x)$ exists for indefinitely large values of n, we infer the existence of a limit function f_0, bounded, B, and equicontinuous $\eta(\epsilon)$, such that $S_\lambda(f_0) = 0$ for a value of λ between 0 and λ_0. Furthermore since the approximating functions are bounded and equicontinuous, it follows that $\int_0^1 f_0^2 ds = 1$, on account of the property $\sum x_i^2 = n-1$, which holds at every stage. This is in contradiction with the hypotheses of the theorem.

Hence, by Theorem VIII, we have for some set x_i

$$g(s) = \pi(s, x) + \lambda_0 Qf$$

for $s = s_i$ $(i = 1, 2, \ldots, n)$. The form of this equation shows that $\pi(s, x)$ is bounded, $2B$, and continuous, $2\eta(\epsilon)$. Consequently, as n increases without limit, there exists a set of polygonal functions $\pi(s, x)$ approaching a limit function f_0 uniformly, where f_0 is bounded, $2B$, and continuous, $2\eta(\epsilon)$. The same limit is approached by $f(s, x)$ of course, and f_0 will lie on H_f. We have then $g = S_{\lambda_0} f_0$, and the theorem is proved.

It should be remarked that λ need not be positive in the above reasoning, and that an inverse direction exists for any value of λ between the largest negative and smallest positive values of λ for which $Sf = 0$ for some function of R_f'.

In the linear case the field of functions on which the transformation operates extends at once to all continuous functions. Also the factor ρ can be suppressed since for invariant directions we can replace f/ρ by f. The above result evidently includes what is essential for the solution of integral equations of the Fredholm type, at least in the case of symmetric kernels.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.